Advances in Visibility Modelling in Urban Environments to support Location Based Services

---

A thesis
submitted in partial fulfilment
of the requirements for the Degree
of
Doctor of Philosophy
in the
University of Canterbury
by
Phil Bartie

University of Canterbury
Christchurch, New Zealand

2011

Acknowledgements

There are a large number of people to thank! Of course firstly Simon, Femke, and Steven for your expert advice, invaluable guidance, constant encouragement and for taking me on. Also to Clive Sabel, Burn Hockey and Mark Billinghurst for getting me started.

Mark Kumler and Eliseo Clementini for agreeing to co-author papers with me from out of the blue. Peter Woodrow from the City Council for granting access to the Christchurch LiDAR data. The University of Canterbury for funding, and the Geospatial Research Centre staff and fellow students for making spatial research so much fun. Special mention must go to Dan, Meghan, Matt and Welshy for our GIS conversations over a cuppa, also Dave, Offy and Eli for chats about coding over a beer, and William Mackaness for ongoing spatial guidance from afar.

Finally to my family in Guernsey and Hsin for their endless love, patience and support.

Table of Contents

Abstract **1**

Chapter 1: Introduction **2**

1.1 Introduction 2

1.2 Thesis Structure 3

1.3 Academic Papers 4

1.4 Additional Contributions by the Candidate 6

1.5 Background 7

1.5.1 Location Based Services 7

1.5.2 Positioning techniques 8

1.5.3 User Interfaces 10

1.5.4 Visibility modelling 11

1.6 Contribution of Thesis 13

Chapter 2: An Egocentric Urban Viewshed: A Method for Landmark Visibility Mapping for Pedestrian Location Based Services **15**

2.1 Summary 15

2.2 Visibility Modelling and Pedestrian Navigation 15

2.3 Visibility Analysis 16

2.3.1 The Urban Cumulative Viewshed 17

2.4 Calculating Egocentric Visibility for LBS 18

2.5 Visibility Implementation for LBS 19

2.5.1 The Database Model 20

2.5.2 Line of Sight Metrics 21

2.5.3 Visibility at Feature Level 23

2.6 Implementation and Evaluation 25

2.6.1 Additional Analysis 28

2.6.2 Mapping the Visibility Metrics 31

2.6.3 Using the Visibility Metrics 32

2.7 Conclusion and Future Work 32

2.8 Acknowledgements 33

Chapter 3: Advancing Visibility Modelling Algorithms for Urban Environments **34**

3.1 Summary 34

3.2 Introduction 34

3.3 Background 36

3.4 Visibility Modelling in Urban Regions 37

3.5 Extensions to Existing Visibility Models 41

3.5.1 Calculating the Field of View Metric 41
3.5.2 Calculating the Visible Area Metric 41
3.5.3 Introducing Distance Factors into Perceived Area Calculations 43
3.5.4 Calculating the Clearness Index for a FOI 44
3.5.5 Calculating how much of a FOI is on the Skyline 45
3.5.6 Pseudo-code 46
3.6 Implementation and Evaluation 48
3.6.1 Synthetic Surface Examples 48
3.7 Real World Examples 50
3.7.1 Visibility Calculation Trials in an Urban Region 50
3.7.2 Studying the Change in Metric Values for a FOI from a New Viewpoint – Case Study 1 54
3.7.3 Studying the Change in Metric Values for a FOI from a New Viewpoint – Case Study 2 56
3.7.4 Mapping Visual Metrics 57
3.8 Conclusion and Future Work 59
3.9 Acknowledgments 61

Chapter 4: Incorporating Vegetation into Visual Exposure Modelling in Urban Environments 62
4.1 Summary 62
4.2 Introduction 62
4.3 Visibility Modelling Background 63
4.3.1 Visibility Modelling Incorporating Vegetation Background 64
4.3.2 Urban Visual Exposure Metrics Background 65
4.4 Accommodating Vegetation within a New Visual Exposure Model 66
4.4.1 Building a Vegetation Map for use in Visibility Studies 67
4.4.2 Implementation 69
4.4.3 Modification to the Line of Sight model to accommodate Vegetation Information 72
4.5 Examples of Visibility Modelling with Vegetation 77
4.5.1 Differences between Visibility modelling with and without vegetation map 77
4.5.2 Taking Account of Seasonal Vegetation Changes 79
4.5.3 Modelling Visibility under Bridges, and Overpasses 81
4.5.4 Reporting the Vegetation Distribution across a Feature 82
4.6 Conclusions and Future Work 82
4.7 Acknowledgments: 84

Chapter 5: Route Ahead Visibility Mapping: A method to model how far ahead a motorist may view a designated route 85
5.1 Summary 85
5.2 Introduction 85
5.3 Preparing the Digital Surface Model 86
5.4 Visibility Analysis 89
5.5 Results 92
5.6 Conclusion 95
5.7 Acknowledgements 96

Chapter 6: A Qualitative Model for Describing the Arrangement of Visible Cityscape Objects from an Egocentric Viewpoint 97
6.1 Summary 97
6.2 Introduction 97
6.3 Spatial Reasoning Review 99
6.3.1 Projective Spatial Reasoning 100
6.3.2 Quaternary Relational Model in R^3 102
6.4 A Combined Model, with Extensions 102
6.4.1 Graded Cases for Y axis 104
6.4.2 Graded Cases for X axis 106
6.4.3 Graded Cases for Z axis 108
6.4.4 Reasoning Summary – Tree form 109
6.4.5 Implementation of Freeman's Primitives for 2.5D LBS 111
6.5 Application of Combined Model Projective Reasoning 114
6.5.1 Visibility Analysis 114
6.5.2 Case Study 1 114
6.5.3 Case Study 2 115
6.5.4 Case Study 3 – Spatial Queries 118
6.6 Conclusion and Future Work 119
6.6.1 Future Work 120

Chapter 7: Referring Expressions in Location Based Services: The case of the 'Opposite' Relation 121
7.1 Summary 121
7.2 Introduction 121
7.3 Background 122
7.3.1 Positional information 123
7.3.2 Visibility Modelling 123
7.3.3 Egocentric Spatial Model 124

7.4 The ‘Opposite’ relation 125
7.4.1 One Dimensional Common Feature 125
7.4.2 Two Dimensional Common Features 128
7.4.3 Additional Considerations for Usage 129
7.5 Conclusions and Future Work 130

Chapter 8: A Combined GIS and Stereo Vision approach to Identifying Building Pixels in Images to Determine Appropriate Colour Terms 132
8.1 Summary 132
8.2 Introduction 132
8.3 Stereo Depth Mapping for Façade Colour Retrieval 134
8.3.1 Structure from Motion Approach 135
8.3.2 Stereo Baseline Vision 137
8.4 Processing Pipeline 140
8.4.1 Sampling Strategy 142
8.4.2 Improving System Robustness 143
8.4.3 Identifying Property Boundary Type 144
8.4.4 Windows 147
8.4.5 Shadows 147
8.5 Colour terms 148
8.5.1 Colour Entropy 149
8.6 Trials and Results 150
8.7 Conclusion and Future Work 152

Chapter 9: Referring Expressions in Urban Environments: Generating and Comprehending Expressions using Bertin’s Visual Variables 155
9.1 Summary 155
9.2 Introduction 155
9.3 Background 157
9.4 Fuzzy Classification for Bertin’s Variables 160
9.4.1 Positional information 161
9.4.2 Orientation 162
9.4.3 Size 164
9.4.4 Colour 165
9.4.5 Texture 166
9.4.6 Form 167
9.5 Referring Expression Generation 169
9.5.1 Most Salient Object - Worked Example 170
9.5.2 Best Referring Expression for Nominated FOI – Worked Example 174

9.6 Description Comprehension 175
9.7 Conclusions and Future Work 177

Chapter 10: Conclusion 179
10.1 Summary of Research 179
10.2 Summary of Outcomes and Conclusions 180
10.3 Current Limitations and Further Work 181

Appendix A: Media Mapping: Using Georeferenced Images and Audio to provide supporting information for the Analysis of Environmental Sensor Datasets. 183
A.1 Summary 183
A.2 Introduction 183
A.3 Background 184
A.4 Application Development 187
A.5 Mobile Data Capture Application 187
A.6 Data Analysis and Playback 189
A.6.1 Mapping 190
A.6.2 Audio 190
A.6.3 Street Images 191
A.6.4 Charting 191
A.7 User Interaction 192
A.8 Examples 193
A.9 Conclusion 195
A.10 Acknowledgements 196

Abstract

People describe and explore space with a strong emphasis on the visual senses, yet modelling the field of view has received little attention within the realm of Location Based Services (LBS), in part due to the lack of useful data. Advances in data capture, such as Light Detection and Ranging (LiDAR), provide new opportunities to build digital city models and expand the range of applications which use visibility analysis. This thesis capitalises on these advances with the development of a visibility model to support a number of innovative LBS functions in an urban region and particular focus is given to the visibility model's supporting role in the formation of referring expressions, the descriptive phrases used to identify objects in a scene, which are relevant when delivering spatial information to the user through a speech based interface. Speech interfaces are particularly useful to mobile users with restricted screen viewing opportunities, such as navigational support for motorists and a wider range of tasks including delivering information to urban pedestrians. As speech recognition accuracies improve so new interaction opportunities will allow users to relate to their surroundings and retrieve information on buildings in view through spoken descriptions. The papers presented in this thesis work towards this goal, by translating spatial information into a form which matches the user's perspective and can be delivered over a speech interface. The foundation is the development of a new visual exposure model for use in urban areas, able to calculate a number of metrics about Features of Interest (FOIs), including the façade area visible and the percentage on the skyline. The impact of urban vegetation as a semi-permeable visual barrier is also considered, and how visual exposure calculations may be adjusted to accommodate under canopy and through canopy views. The model may be used by pedestrian LBSs, or applied to vehicle navigation tasks to determine how much of a route ahead is in view for a car driver, identifying the sections with limited visibility or the best places for an overtaking manoeuvre. Delivering information via a speech interface requires FOI positions to be defined according to projective space relating to the user's viewpoint, rather than topological or metric space, and this is handled using a new egocentric model. Finally descriptions of the FOIs are considered, including a method to automatically collect façade colours by excluding foreground objects, and a model to determine the most appropriate description to direct the LBS user's attention to a FOI in view.

Chapter 1: Introduction

1.1 Introduction

The "information age" is a time of unequalled access to knowledge, with capabilities to transfer information more freely than ever before. However "information overload" can become a problem as the quantity of data increases, requiring better solutions to sustain search efficiency and maintain information relevance (Morville 2005). One strategy to improve search effectiveness is to automatically include location as a criterion, thereby limiting results by spatial significance. Location awareness is particularly relevant for the post-desktop era of ubiquitous computing (Weiser 1993, Weiser *et al.* 1999), especially for mobile devices referred to as Location Based Services (LBS), a good account of which can be found in Jiang and Yao (2006).

Location is most typically introduced as a search parameter based on proximity, filtering the results based on an arbitrary distance measured in Euclidean or network space (Mountain and MacFarlane 2007). However people describe and explore space with a heavy emphasis on the visual senses, something under-used as a search parameter within LBSs (May *et al.* 2005). In part this has been due to the lack of useful data but recent advances in data capture, such as Light Detection and Ranging (LiDAR), provide an economically viable way to build digital city models leading the way to the inclusion of visibility analysis in a wider range of urban applications (Palmer and Shan 2002, Rottensteiner and Briese 2002, Bartie and Mackaness 2006).

This thesis develops visibility modelling and its inclusion in a range of novel functions for use in urban LBS with the goal of supporting new types of user interaction. This begins with the development of a new model for filtering and quantifying the visibility of features of interest (FOI) taking into account views through and under vegetation canopy. A model able to store the relationship for the results from an egocentric perspective is developed, matching the user's view, providing the foundation for describing visible features using projective spatial relationship terms, such as "in front of" and "left of" (Freeman 1975, Clementini and Billen 2006, Clementini 2009). This can be combined with other descriptive attributes, such as building colour, to form useful referring expressions (Dale and Reiter 1995). These referring expressions will be required in advancing speech based user interfaces, which are considered to be a beneficial addition to future LBS devices as they allow the user to maintain focus on their surroundings, operating eyes-free and hands-free (Goose *et al.* 2003, Sharma *et al.* 2003, Howell *et al.* 2005, Bartie and Mackaness 2006).

The aim of the thesis is to advance the modelling and communication of items in the field of view for use in urban LBS, by addressing the following research objectives:

1. Develop a visual exposure model for use in an urban LBS, which delivers a range of new metrics suitable to quantify the visibility of nominated features of interest.

2. Accommodate under canopy and through canopy views of features of interest within the visual exposure model.

3. Express relationships between visible features from the user's viewpoint using an egocentric projective model of space, with support for storage and querying.

4. Determine the most suitable referring expression to describe a visible item to the LBS user, using visual attributes (e.g. size, colour, position).

1.2 Thesis Structure

The thesis begins by examining the wider context of LBS with a review of current positioning solutions and mobile user interfaces. A new visual exposure model is developed in Chapters 2 and 3, suited to urban environments, able to generate a number of metrics for Features of Interest (FOI) in a scene. The term FOI is used throughout, rather than just the usual Points of Interest (POI), as consideration is given to regions of visibility corresponding to urban objects (e.g. statues, buildings). The advances made include measuring visible façade areas, with detailed information on which foreground features block the view leading to the introduction of a *viewing clearness* metric. The impact of vegetation within the visual exposure model is addressed in Chapter 4, such that partial through canopy and under canopy views are modelled. The use of the model as part of a navigation task is explored in Chapter 5, such that a motorist is pre-informed of blind corners and good overtaking opportunities along a defined route.

As speech recognition improves so speech based user interfaces will become a more viable interaction method on mobile devices. The need to verbalise spatial information in a format that can be easily comprehended raises the topic of how referring expressions may be generated for visible entities based on their appearance from the user's perspective. Part of the issue is in how objects are described in egocentric projective space (Chapter 6), but also in how descriptions may be formed using the most salient visual attributes (Chapters 7 & 9). Both of these concepts are discussed, and the issue of how building colour information may be collected is also examined in Chapter 8. Here a new approach is demonstrated to combine computer vision techniques with spatial analysis to automate the collection of house and boundary wall colours, filtering out foreground objects. This information is used in conjunction with other datasets to form useful descriptions of the features in view, such that a LBS user's attention could be directed to a target feature.

A schematic is shown in Figure 1.1 of a LBS application which uses visibility modelling and accesses several data sources to deliver information to a user. The corresponding thesis chapters are denoted in this figure.

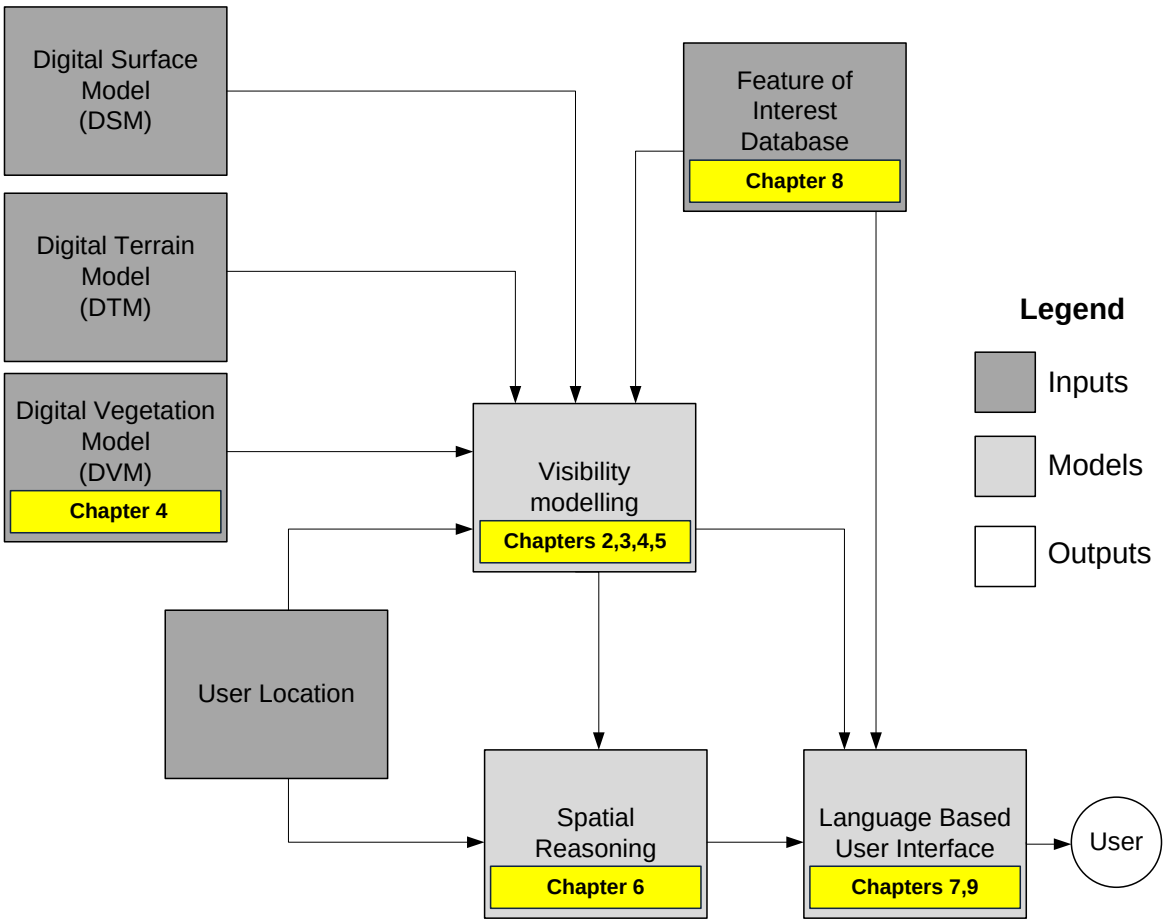


Figure 1.1: Schematic of Information Flow for a LBS with Corresponding Thesis Chapters

1.3 Academic Papers

This doctoral thesis was completed by preparing a series of academic papers, with an overall ambition of advancing the modelling and communication of items in the field of view for use in urban LBSs. In accordance with University of Canterbury standards the candidate is the first author and primary contributor in all papers. The relationship between objectives, papers and chapters is shown next.

Objective 1) Develop a visual exposure model for use in an urban LBS, which delivers a range of new metrics suitable to quantify the visibility of nominated features of interest.

- **Chapter 2)** Bartie, P. J., Mills, S. & Kingham, S., 2008. An egocentric urban viewshed: a method for landmark visibility mapping for pedestrian location based services. *In Moore, A. & Drecki, I. (eds.) Geospatial Vision - New Dimensions in Cartography. New Zealand, Springer, 61-85.*

- **Chapter 3)** Bartie, P. J., Reitsma, F., Kingham, S. & Mills, S., 2010. Advancing visibility modelling algorithms for urban environments. *Computers Environment and Urban Systems, 34(6), 518-531.*

Objective 2) Accommodate under canopy and through canopy views of features of interest within the visual exposure model.

- **Chapter 4)** Bartie, P., Reitsma, F., Kingham, S. & Mills, S., 2011. Incorporating vegetation into visual exposure modelling in urban environments *International Journal of Geographical Information Science, 5 (5), 851-868.*

- **Chapter 5)** Bartie, P. J. & Kumler, M. P., 2010. Route ahead visibility mapping: a method to model how far ahead a motorist may view a designated route. *Journal of Maps, April 2010, 84-95.*

Objective 3) Express relationships between visible features from the user's viewpoint using an egocentric projective model of space, with support for storage and querying.

- **Chapter 6)** Bartie, P., Clementini, E., Reitsma, F. & Kingham, S., *(revise and resubmit).* A qualitative model for describing the arrangement of visible cityscape objects from an egocentric viewpoint. (*revise and resubmit - Geoinformatica*)

Objective 4) Determine the most suitable referring expression to describe a visible item to the LBS user, using visual attributes (e.g. size, colour, position).

- **Chapter 7)** Bartie, P., Reitsma, F. ,Clementini, E., & Kingham, S, 2011. Referring expressions in location based services: the case of the "opposite" relation, in Semantic Enrichment of 3D city models for sustainable urban development, *SECOGIS 2011, Springer-Verlag, Brussels, Belgium.*

- **Chapter 8)** Bartie, P., Mills, S. & Reitsma, F., 2011. Building Colour Terms: a combined GIS and stereo vision approach to identifying building pixels in images to determine appropriate colour terms, *Journal of Spatial Information Science, 2(2011), 59-83.*

- **Chapter 9)** Bartie, P., Reitsma, F., & Kingham, S., *(yet to submit).* Referring expressions in urban environments: generating and comprehending expressions using Bertin's visual variables. *Journal of Location Based Services (to be submitted).*

1.4 Additional Contributions by the Candidate

It should also be noted that during the PhD programme the candidate published two additional papers. The first is included in Appendix A, and describes a technique to use a smartphone to capture video and audio data with the corresponding time and location details, intended for use in conjunction with synchronised environmental sensors. The media is able to assist researchers in the data analysis stage by providing context from the time of collection, which can help to understand variations in the environmental data caused by temporal events, such as an increase in air pollution cause by a passing bus. The second paper relates to the presentation of thermal imaging on a LiDAR sourced city model.

- Bartie, P., & Kingham, S. (2010). Media mapping: using georeferenced images and audio to provide supporting information for the analysis of environmental sensor datasets. *The Journal of the Open Source Geospatial Foundation, 8*.

- O'Donohue, D., Mills, S., Kingham, S., Bartie, P., & Park, D. (2008). *Combined thermal-LiDAR imagery for urban mapping.* Paper presented at the Image and Vision Computing New Zealand, 2008. 23rd International Conference Christchurch, NZ.

1.5 Background

Developments in a number of technologies have created opportunities for new kinds of interaction with computers, and digital data. These include more powerful mobile computing platforms, better location solutions across a wider range of environments, improved user interfaces, and access to data through faster network connections. As a result a new breed of mobile computing exists where users can query their environment *in situ*. The following summarises some of the key research in these areas providing additional background material to that in each of the thesis chapters.

1.5.1 Location Based Services

LBSs are location-aware and context-aware applications running on mobile devices capable of providing information to the user on their surroundings (Jiang and Yao 2006) the majority of which provide way-finding assistance (Douglas *et al.* 2006).

Cyberguide (Long *et al.* 1996) was a LBS pioneer, able to calculate its location indoors using infrared beacons and outdoors using the Global Positioning System (GPS), providing location customised information to tourists. The system demonstrated that mobile computing was able to usefully adapt information delivery based on location and place histories, offering an alternative to a human tour guide.

A wide variety of location aware applications have been developed since, including GUIDE a virtual guidebook (Davies *et al.* 1999, Cheverst *et al.* 2000), EASYGO for public transport information (Gartner *et al.* 2007), GEONOTES for attributing space using virtual tags (Espinoza *et al.* 2001), urban gaming (Benford *et al.* 2006), friend finding (Strassman and Collier 2004, Neer 2011), and way-finding applications (May *et al.* 2003, May *et al.* 2005, Google 2006).

The uptake of LBS by the population has been fairly slow, which can be partially attributed to poor user experiences, a lack of service dependability, and a lack of perceived ownership benefits (Chincholle *et al.* 2002). Furthermore many potential users are concerned about issues of privacy (Duckham and Kulik 2006) and security (Cahill *et al.* 2003). However driven by the inclusion of GPS in smartphones and the availability of free, or very cheap, applications it appears that LBS popularity is on the rise (Williams 2009, Kim 2011). There are also initiatives such as OpenLS (Zipf 2004, OGC 2008), which aim to standardise access protocols to spatial data services and content repositories, facilitating wider adoption across a growing range of devices.

In parallel to the progress of LBS has been the development of mobile Augmented Reality browsers, which superimpose digital data in real-time onto views captured by a device camera, for example Layar (www.layar.com) and Wikitude (www.wikitude.org). These devices require

fairly high location resolution and a digital compass to track device orientation, and are sometimes referred to using the term Mobile Spatial Interaction (Fröhlich *et al.* 2009).

1.5.2 Positioning techniques

A LBS should be able to automatically and continuously determine the user's position so that relevant information may be retrieved and presented. Global Navigation Satellite Systems (GNSS), such as GPS, have become a widely adopted solution for doing this as they offer global coverage with minimal operating and maintenance costs for the end user; a full operational overview can be found in Kaplan and Hegaty (2006). Despite the widespread popularity of GPS it has an "Achilles heel" (Raper *et al.* 2007), struggling to operate in "urban canyons" where the direct lines of sight to the GPS satellites are occluded, and signals reach the receiver after being reflected from nearby surfaces, known as multipath signals. The situation is worse for pedestrians positioned on the pavement, close to tall buildings, as satellites are occluded to a greater extent than for motorists positioned near the centre of the street.

There are several strategies to improve GPS performance, two of the most successful currently implemented are the Satellite-Based Augmentation System (SBAS) (Ott *et al.* 2005) and Assisted-GPS (A-GPS) (Cho 2004). SBAS is a satellite based Differential-GPS (DGPS) system consisting of a separate series of satellites and ground stations which transmit GPS correction details, including data on signal delays due to atmospheric effects and clock drift, typically improving location accuracy for equipped devices to 3 metre accuracy (Garmin 2011). SBAS is known as WAAS in USA, EGNOS in Europe, and MSAS in Japan, however Australia and New Zealand are not currently supported by such a system.

A-GPS aims to decrease the time to first location fix by using a networked service to provide the mobile client with supporting information on relevant satellite orbital data, ionospheric conditions, and assistance in matching fragments of GPS signal received. The benefits are a decrease in the time taken to calculate a location, and the ability to decode weaker signals, but not an improvement in location accuracy. It is available world-wide but requires that the mobile device has an internet connection to download the relevant supplementary data.

In addition to these developments the latest generation of "High Sensitivity GPS" devices are able to resolve locations in urban canyons and albeit at low resolution inside some buildings (Lachapelle 2004). The situation will be further improved as new GNSSs become available (e.g. Galileo) with the increase in satellite numbers and improved signal robustness regarding multipath (Fantino *et al.* 2008) enabling higher location accuracy in the urban environment. Accuracy could also be improved using Real-Time Kinematic (RTK) positioning methods based on carrier-phase measurements, which are readily available in GNSS receivers, locating the phone in real-time to the nearest decimetre (Nokia 2006, Wirola *et al.* 2007).

Position robustness may also be improved by supplementing GNSS solutions with extra sensors known as Inertial Measurement Units (IMUs) able to supply relative motion changes. These can be placed on the user's foot to exploit human walking kinematics and therefore to estimate the user's position in between GNSS updates (Radoczky 2007). An IMU incorporates a number of sensors which record movement direction and magnitude, including a combination of accelerometers able to measure movement forces in 3 dimensions, gyroscopes able to track direction changes, and barometers to measure elevation changes, which is particularly useful inside buildings for floor level estimation.

IMUs are able to assist GNSS in challenging environments (e.g. indoors, urban canyons) but without GNSS updates the positional errors accumulate over time and positional drift occurs. The rate of this drift varies depending on the quality of the sensors in the IMU, but can be improved by monitoring the user's steps and performing zero velocity updates. These are when the IMU is known to be static as the user's foot is placed on the ground, and therefore the sensors may be re-calibrated (Mezentsev *et al.* 2005). The GNSS and IMU sensor inputs are integrated into a single best estimate of location using a Kalman filter (Yang *et al.* 2008), which has demonstrated to be effective in a number of projects (Godha *et al.* 2005, Ott *et al.* 2005).

There are a number of other emerging technologies in various stages of development which could one day provide a high resolution urban positioning solution. For example terrestrial radio waves fill urban environments, and Robust Surface Navigation (RNS) (Hodge 2007) relies on using these "signals of opportunity" to calculate the user's location. Signals suitable for this include FM Radio (Krumm *et al.* 2003), AM signals (McEllroy *et al.* 2007), cellular phone network signals (Zhou *et al.* 2005, Chen *et al.* 2006) and other shorter range signals such as Wi-Fi (Hightower *et al.* 2005).

Dense networks of Wi-Fi base stations found in cities provide a useful fast positioning solution. Wi-Fi equipped mobile platforms permit developers to access signal strength information from all the surrounding access points, meaning position triangulation is possible. Trials have shown that a median location accuracy of 30m can be achieved in urban areas (LaMarca *et al.* 2005). This method requires a database of the physical locations of each Wi-Fi base station with its corresponding unique Media Access Control (MAC) address. There are a number of open source databases, maintained by community volunteers, which hold this information, such as Placelab (LaMarca *et al.* 2005). However Wi-Fi signals may only cover a small proportion of the city, and as the base stations are not controlled by a single operator any unit may be moved or turned off without prior notice. A number of commercial companies have begun to offer applications which combine the location benefits from cell tower, Wi-Fi and GPS positioning solutions, such as Skyhook (Skyhook 2008).

Even greater positional accuracy is possible using scene-analysis methods, whereby the user captures an image of the surrounding environment using a mobile phone camera, then sends the

image to a server which geometrically straightens it for comparison against a database of images from known locations. Typically the key points identified in each image include corners of window frames or doors, which change infrequently. The result is that the user's orientation and location can be calculated, usually to within one metre (Randerson 2004). However this technique requires considerable computational power on the server side and frequent image database updates. A manual equivalent exists whereby the user specifies which landmarks are visible, and the mobile device calculates the possible locations that they may be (Kray and Kortuem 2004). A negative aspect of both of these methods is that the system latency renders the techniques unsuitable for continuous tracking.

This tracking issue is also of interest in robotics research, referred to by the acronym SLAM which means Simultaneous Localisation and Mapping. Until recently expensive sensor arrays which included laser and IMUs were required to track robot movements over great distances, but new research has shown that a single digital camera can be used to perform localisation over a 2.5km walk (Botterill *et al.* 2010). Developments such as these should lead to sustained sub-metre accurate positioning for LBS in urban environments in the near future.

1.5.3 User Interfaces

Interface design must put the user centre stage and consider not only the task requirements but also the environment in which the user and computer connect (Worboys 2001). A very successful interface is one in which the user is not aware of its existence (Weiser 1994), or may only require continuous partial attention rather than be totally engaging (Weiser and Brown 1996). A variety of interfaces, reviewed below, can support this spectrum of interaction.

Graphical user interfaces in mobile devices must convey essential information via limited screen space in difficult environmental conditions (e.g. bright sunlight) and use input methods suitable while moving (Dillemuth 2005). LBSs are able to overcome screen size issues to some extent by carrying out map panning, rotation, and zooming automatically based on user context and location (Frank *et al.* 2004), and by presenting simplified representations of the surrounding spatial information using a sketch map approach (Richter *et al.* 2008, Schmid *et al.* 2010). However graphical user interfaces remain a distraction, requiring the user to refocus attention away from their surroundings.

It has been shown that users are able to carry out multiple tasks more easily if they are using a number of sensory modalities (Allport *et al.* 1972). One way to attract attention at a specific time is to use haptic interfaces, which make use of the sense of touch. Research has shown that it is possible to guide someone using an array of vibration devices mounted in a vest (Ertan *et al.* 1998), or shoes (Frey 2007), to indicate when and where to turn. Similarly a mobile phone in a pocket could vibrate to indicate that the user should turn at a junction, and basic interaction may

take place by tapping the device, measured by internal accelerometers, for example to repeat an audio instruction.

A fuller interaction experience is possible using a speech based interface, which is particularly suited to LBS as it offers discreet hands-free and eyes-free use, enabling the user to keep focussed on their environment while receiving (speech synthesis) and giving instructions (speech recognition).

Speech recognition accuracy is dependent on a number of factors including system training and background noise. Noise cancelling microphones are able to assist to some degree, but in many cases speech recognition performs best when limited to a small vocabulary of command words. Demonstrations of speech interfaces for information retrieval can be found in a variety of fields, including emergency response centres, database querying, and searching technical manuals (Goose *et al.* 2003, Sharma *et al.* 2003, Du and Crestani 2004). A number of the advantages and disadvantages of speech user interfaces are given in Table 1-1.

Table 1-1: Advantages and Disadvantages of Speech User Interfaces for LBS

Advantages • Low power consumption compared to LCD screen • Natural conversational communication • No distraction from viewing the surroundings (hands free, eyes free) • Accessible in rain without risk of damage to device • Accessible to visually impaired people • Familiarity with the interface hardware, lower learning curve • Lightweight hardware (headphones, microphone), inexpensive – unlike head mounted displays • Compact, yet without the constraints of limited screen area and map design • Secure and discreet – the user may not want to be seen looking at maps or appearing lost
Disadvantages • Speech recognition errors in noisy streets • User's accent and speed of speaking affects accuracy of voice recognition (system coaching required) • Does not allow a user to browse the information • Confusion of target identification and selection in crowded regions • Cannot be used by hearing impaired

For LBSs to support natural interactions through speech interfaces, leading to more intelligent services, there is a requirement to more closely mimic what a co-located human would communicate based on a shared experience of the surrounding space (Winter and Wu 2009). This includes the ability to refer to surrounding features, and model which items are in view. The next section gives a brief overview of visibility modelling, and its use to establish which features can be observed.

1.5.4 Visibility modelling

Visibility modelling can be used to determine which regions may be viewed from a given observation point. In urban areas this has tended to be based on isovist theory (Tandy 1967, Benedikt 1979) using building plans to determine the limits of view, without consideration for

topography or building height. Recently 3D isovist modelling has become a possibility by accessing city elevation models which include the urban form (Morello and Ratti 2009).

In rural regions viewsheds are used to determine what is visible from a vantage point by calculating which cells of an elevation model are visible to an observer (Tandy 1967, Lynch 1976), using a line of sight algorithm (Fisher 1993). Elevation data may be stored in rasters or Triangular Irregular Networks (De Floriani and Magillo 1994, De Floriani *et al.* 1997), and uncertainty may be reflected in the results (Fisher 1994b, Fisher 1995). They have been used in a wide range of applications including finding the most hidden route or those with the best views (Lee and Stucky 1998), studying the visual impact of wind farms (Kidner *et al.* 1999), and in landscape planning (Fisher 1995). Performance has also received a lot of attention as processing can be computationally intensive, especially across high resolution elevation surfaces (De Floriani *et al.* 2000, Rana and Morley 2002, Rana 2003, Ying *et al.* 2006).

The results from viewshed analysis describes "vista space" (Montello 1993), that is the space which can be seen from a static location with only movements of the observer's head. Figure 1.2 illustrates its fragmented nature in comparison to Euclidean space and network space.

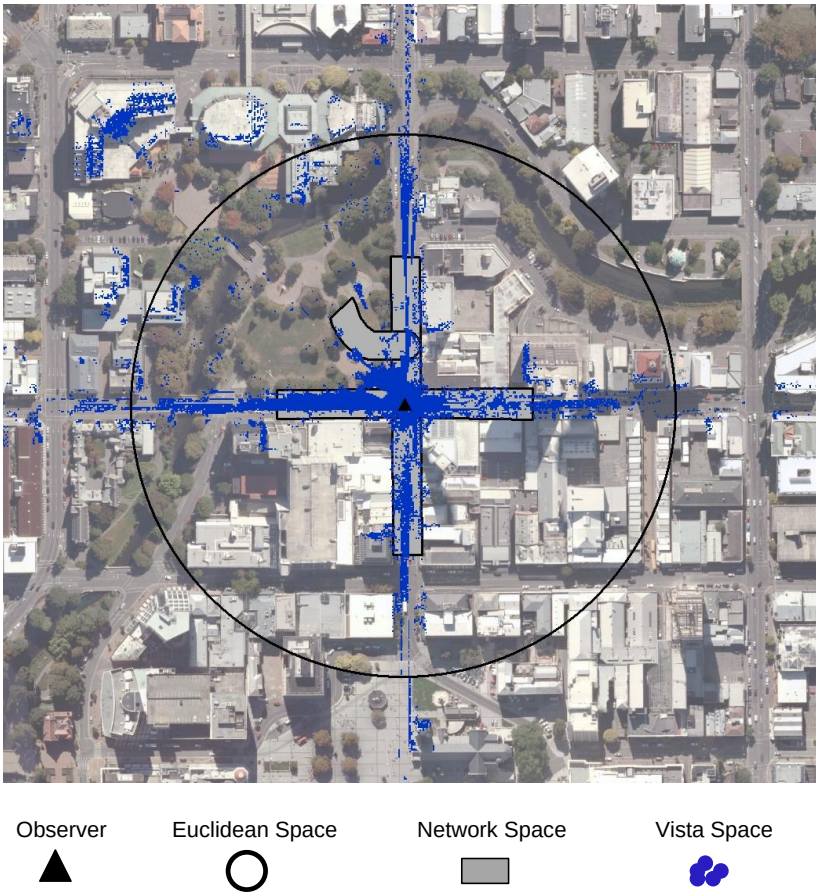


Figure 1.2: Spatial filter examples: Euclidean space (200m radius), Network space (100m), and Vista Space

Vista space is relevant to a LBS user as it defines the current surroundings, through the objects which are visible, and may be included as a filtering method. It is also appropriate for Mobile AR applications which currently filter information using keyword and proximity leading to screen cluttered with labels referring to items out of sight. This will become more of a problem as the quantity of georeferenced data increases, particularly as crowd sourced volunteered information becomes more commonplace (Black and Coast 2007, Goodchild 2007). The introduction of a visibility model to determine which items are currently in view would allow information to be filtered in accordance with the user's real world experience, improve the user experience, and pave the way for new user interface interactions. There are only a few examples of visibility filtering in LBS (Beeharee and Steed 2005, Bartie and Mackaness 2006, Maierhofer *et al.* 2007, Carswell *et al.* 2010). These studies have found visibility to be a useful egocentric filter, but have failed to use the results to describe object visibility in any great detail.

1.6 Contribution of Thesis

This thesis contributes to two main areas of LBS research. These are the advancement of visual exposure modelling in urban regions, and the preparation of information for delivery through a speech interface.

The advancements in visual exposure modelling provide facilities to describe FOIs according to a range of new metrics in urban environment using a high resolution city model (Chapters 2,3). In addition modelling the line of sight through vegetation canopy is addressed by combining a DSM, a DTM and multiband imagery (Chapter 4). While vegetation has been included in visibility modelling to some degree before (Dean 1997, Llobera 2007a), its effect on visual exposure has not previously been researched. In addition the raster vegetation model introduced here can accommodate high resolution density variations across the canopy

Information delivery through a speech interface requires translation into natural language phrases. To assist the user relationships between visible objects are referred to using projective spatial relations, matching the user's viewpoint. Chapter 6 presents a new projective model which combines and extends existing models allowing for projective relationships between buildings to be serialised for storage and querying purposes. Chapters 7 and 9 deal with referring expressions, and how rare or unique visual characteristics can be used to disambiguate items in view. In support of this Chapter 8 develops an automated technique to collect colour information terms from building façades by filtering out foreground objects using a combined computer vision and GIS approach. This information can be included in a building's description when forming the referring expressions.

Another overall contribution is in establishing visibility modelling in a range of spatial roles, functions, and definitions. Chapter 5 demonstrates how visibility modelling can be used in a car

navigation application for estimating how far along a route a driver can see, thereby giving advanced warning of sections with limited visibility or good overtaking opportunities. Chapter 7 shows that a simple concept such as “spatially opposite” relies on a mutual visibility between two reference objects and a common central feature.

The next three chapters develop visual exposure modelling in urban environments.

Chapter 2: An Egocentric Urban Viewshed: A Method for Landmark Visibility Mapping for Pedestrian Location Based Services

2.1 Summary

A variety of information can be provided to pedestrians using location based services in support of tasks such as way-finding. Typically current location aware systems use proximity to filter databases for contextual information. We show that a filter based on the visibility of features is a useful additional capability made possible through the use of digital surface models. A number of visibility metrics are suggested for adoption by a location based service, to provide quantitative visibility information so that items of interest may be ranked according to a meaningful priority. Real world experiences validate the usefulness of these metrics, and a number of improvements are suggested.

2.2 Visibility Modelling and Pedestrian Navigation

There is a growing interest in the development of location based services (LBS) in support of pedestrian activities, both rural and urban (Jiang and Yao 2006). The research presented in this chapter is in anticipation of devices able to support natural pedestrian way-finding and navigation in an urban context. It is argued that the requirements for pedestrian navigation are quite different from that of vehicle navigation. While junctions form a key navigation component for motorists, pedestrians more often use landmarks as cues (Millonig and Schechtner 2007). The urban environment is defined by these landmarks, their organisation, and interrelationships (Fisher-Gewirtzman and Wagner 2003) and there is a strong linkage between what a pedestrian can see and how they comprehend a city. We therefore argue that integrating the capability to use landmark visibility information in a navigational device for urban pedestrian use, requires 'egocentric visibility modelling'.

The term 'viewshed' has existed in landscape architecture since the 1960s (Tandy 1967, Lynch 1976), and has been adopted by many disciplines. A viewshed depicts areas which can be seen from a designated observation point, generated by calculating lines-of-sight (LOS) from that point to all other locations within the study area. Most Geographic Information Systems (GIS) offer the functionality to carry out such calculations (De Smith *et al.* 2007). Visibility is one of the most commonly used GIS analysis tools (Davidson *et al.* 1993) with an extensive catalogue of research work, including siting radio masts (De Floriani *et al.* 1994a), locating the most scenic or most hidden routes (Stucky 1998), landscape planning (Fisher 1996), as a weapon

surrogate in military exercises (Baer *et al.* 2005), and in examining spatial openness in built environments (Fisher-Gewirtzman and Wagner 2003).

Visibility studies within GIS require access to a digital terrain dataset for the area of interest. These are usually raster grid datasets, considered to be 2.5 dimensional, recording a single elevation value (z) for any location (x,y). In general terms these are known as Digital Elevation Models (DEM), but may be more specifically referred to as Digital Terrain Models (DTM) if they reflect the elevation values of the bare earth, or Digital Surface Models (DSM) if they capture building and vegetation elevations.

There have been numerous previous studies on urban visibility, these have tended to use the 2 dimensional boundary of buildings to calculate isovists (Tandy 1967, Benedikt 1979, Turner *et al.* 2001). To more closely model urban visibility it is necessary to source a DSM at high resolution, such that building and vegetation profiles are captured accurately. Light Detection and Ranging (LiDAR) remote data capture techniques have been shown to be suitable for this in urban studies (Palmer and Shan 2002, Rottensteiner and Briese 2002), and are considered superior for LOS calculations than using community contributed 3D models of inconsistent and questionable accuracy, such as with the Google 3D Warehouse.

This chapter discusses visibility modelling techniques, and presents a number of additional metrics that will allow future location based services to report landmark information in a more intuitive manner, facilitating the exploration of the city. The chapter draws attention to a number of relevant areas of visibility research, including cumulative visibility. It then explains the line-of-sight algorithm used in this research, and a supporting database architecture to model Features of Interest (FOI). Finally the method is demonstrated in a real world situation.

2.3 Visibility Analysis

If every terrain cell in a line-of-sight path is considered between an observer and target it is referred to as the ‘golden case’ (Rana and Morley 2002). Although providing the most accurate results from a terrain model this method is computationally expensive, and therefore much of the previous research has focussed on techniques to reduce the number of calculations by considering only visually important cells. Examples of this include using Triangulated Irregular Networks (TINs) (De Floriani and Magillo 1994), or filtering based on topographic features (Rana and Morley 2002). These essentially look to simplify the terrain complexity, or reduce the number of observer-target pairs considered in viewshed generation.

The ‘golden case’ may be maintained whilst offering rapid retrieval of visibility details by using a Complete Intervisibility Database (CID) (Caldwell *et al.* 2003), also referred to as a visibility graph (O'Sullivan and Turner 2001), or visibility matrix (Puppo and Marzano 1997). These approaches store the pre-calculated viewshed results from every possible location in a

study region, meaning future users only require a simple lookup to return the stored viewshed result. The computational cost of producing a CID is very high, being an $O(n^2)$ calculation, therefore parallel or grid processing techniques are often used (Llobera *et al.* 2004).

By storing the complete set of viewsheds a number of additional attributes are available for analysis, such as the Cumulative Visibility (Wheatley 1995), or Visual Magnitude (Llobera 2003). These depict the total number of times a cell can be viewed from elsewhere, indicating highly visible regions. Topographically prominent areas such as ridges and peaks often feature highly, however visually prominent landscapes may not necessarily be topographically prominent, such as the high intervisibility which occurs in valleys (Llobera 2003).

2.3.1 The Urban Cumulative Viewshed

In the context of rural DTMs, peaks and ridgelines form important spaces which act as barriers, and occupiable vantage points. In urban DSMs these ridgelines correlate to building roofs, and are not generally occupiable. This has a number of implications with regard to previous research focussed on reducing observer-target pairs, which consider ridgelines to be significant vantage points (e.g. Rana and Morley 2002). To faithfully replicate the situation in urban space, a cumulative visibility map must be based on a DSM and restrict observer locations to those areas accessible by pedestrians.

Figure 2.1 shows cumulative visibility for a section of Christchurch, New Zealand. This was produced by calculating viewsheds from 20,000 locations selected at random from all publicly accessible spaces, and summing the results to indicate how many times a cell can be seen. The DSM was created at 1 metre resolution from LiDAR return information. To reduce the impact of edge-effects the analysis was carried out over a larger extent than that shown (Llobera 2003).

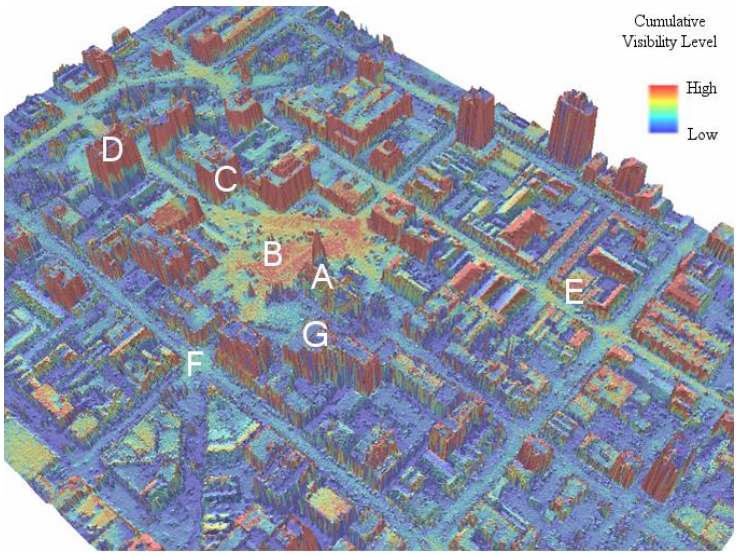


Figure 2.1: Cumulative Visibility Map for Christchurch, NZ

A number of observations can be made:

- Tall buildings have high cumulative visibility values (e.g. Christchurch Cathedral - A)
- Open spaces have high cumulative visibility (e.g. Cathedral Square - B)
- The building frontages adjacent to streets have high visibility values for the entire face (e.g. building C) whilst those faces surrounded by low rise development only receive high scores for the uppermost sections of the face (e.g. building D)
- Low rise buildings only visible from a single street receive mid-range scores (e.g. E)
- Street intersections receive a high score, as they can be viewed from a number of directions (e.g. F)
- The roofs of buildings have low scores as they cannot be viewed from street level (e.g. G).

Depending on the intended purpose, the computational cost of cumulative viewsheds may render them inappropriate. De Floriani (1994a) suggested that the cost will not be repaid if there are a minimal number of observation locations. In terms of an LBS where viewshed information is only required for a single user location, it would therefore follow that a cumulative viewshed would not be appropriate. However the cumulative viewshed indicates which areas in an urban scene are important for visibility analysis, and this information could be used in a strategy to reduce the number of target locations for real time visibility analysis conducted on a mobile device.

2.4 Calculating Egocentric Visibility for LBS

Whilst GISs typically consider the world from above with all areas equally important, an LBS takes an egocentric viewpoint (Meng 2005, Reichenbacher 2005). A LBS is defined by Jiang and Yao (2006) as an application which is both location-aware, and context-aware, therefore requiring information on the user's position and surroundings. Currently LBSs use proximity as a spatial filter to retrieve relevant contextual information, a notable exception is the Edinburgh Augmented Reality System (EARS) (Bartie and Mackaness 2006) which is able to filter information based on the visibility of FOIs from the user's location.

EARS accesses a database of pre-calculated visibility results for 86 FOIs located around the city of Edinburgh. These results were calculated using ESRI ArcInfo and stored in a relational database management system for rapid sub-second retrieval while on location. As the user explores the city, Global Positioning System (GPS) values are used to locate the user and the application reports what can be seen from the current location. There are a number of drawbacks to this approach including:

- the user's height is fixed to 1.74m for all visibility calculations
- minimal quantitative information is available regarding feature visibility
- the system is unable to accommodate user or community contributed FOIs, as viewshed functionality is not provided on board
- any updates to the DSM require all the visibility calculations to be re-run.

An LBS able to provide the user with information about the visibility of FOIs would be able to guide the user by referring to landmarks (Michon and Denis 2001, Raubal and Winter 2002, Goodman *et al.* 2004, Ross *et al.* 2004, May *et al.* 2005, Millonig and Schechtner 2007), or inform the user about the current surroundings in a natural way. It is therefore necessary to provide an LOS algorithm for use in a LBS which can provide real time quantitative information on the visibility of FOIs.

2.5 Visibility Implementation for LBS

The visibility algorithms implemented in GISs are often the subject of debate; Fisher (1991) reported that different packages gave significantly different results. Source code for Open Source GIS applications are in the public domain available for scrutiny, whilst the algorithm implementations of commercial software is unknown. A useful survey of the visibility functionality in GIS can be found in a publicly available report to the US Army Line of Sight Technical Working Group (US Army Corps of Engineers 2004).

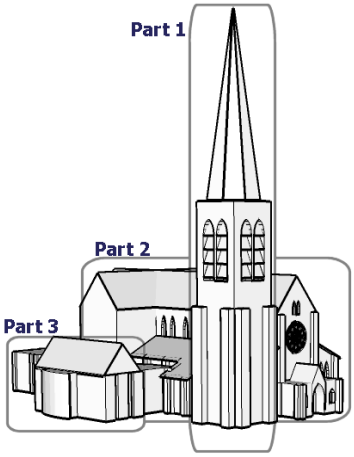
Riggs and Dean (2007) showed through field trials that predicted viewsheds and surveyed results had lower discrepancies when using higher resolution DSMs. For urban studies using LiDAR datasets it is hoped that discrepancies will be small, although it is acknowledged that a DSM is a 2.5D dataset, and will not report true visibility values under bridges, overpasses, or under vegetation canopy.

Llobera (2003) introduced the concept of 'visual exposure', and suggested that this dynamic aspect of visibility had been overlooked within previous research. Visual exposure focuses on how much of a feature can be viewed from the surrounding space, enabling the creation of surfaces to show in which direction a viewer would need to move to view the target more, or less, clearly. This technique can be used to find visual corridors, or visual ridges, and forms a useful basis for considering LOS in the context of LBS.

For this research a toolkit was written to allow experimentation with the 'golden case' LOS algorithm. The source datasets were a surface model, an observer location, and a database of feature locations. For this study the test area selected was the city of Christchurch, New Zealand. All calculations were carried out using New Zealand Map Grid (NZMG), with the facility for a user height to be specified. LiDAR data was sourced from the Christchurch City Council, and a DSM rendered at 1 metre resolution. For simplicity vegetation was treated as a visual barrier, although the concept of partially obscured views through vegetation has been examined (Llobera 2007a).

2.5.1 The Database Model

A database model was designed such that each FOI entity could be divided into component parts. For example Christchurch Cathedral (Figure 2.2) could be divided into three parts to represent the spire, the main building, and café annex. Each FOI part could be assigned a number of target locations so the visually important aspects of the structure could be explicitly modelled. It is necessary to place targets around the base of FOIs so that as higher targets are obscured by the building's walls on approach, the LBS does not consider the FOI to be out of view (Figure 2.3). It therefore follows that a greater proportion of the targets on an FOI should be visible as the observer moves away, although the target will occupy a smaller part of the field of view.



Part 1 – Spire
Part 2 – Main body
Part 3 – Café

Figure 2.2: Christchurch Cathedral, NZ (3D model by ZNO, sourced from Google 3D Warehouse)

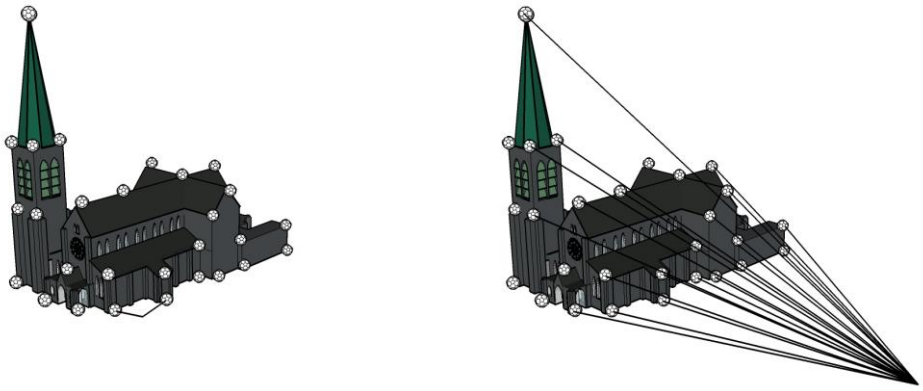


Figure 2.3: Target Placement (3D model by ZNO, sourced from Google 3D Warehouse)

Provision was made within the database structure (Figure 2.4) to scan targets according to a visual significance hierarchy. This allowed the LBS to perform a number of scans at varying target densities, firstly to detect the visibility of FOIs, secondly to quantify the visibility information. For this research performance was not a primary consideration, therefore all 2568 cells within the building boundary were used as target locations, at a resolution of 1 target per square metre.

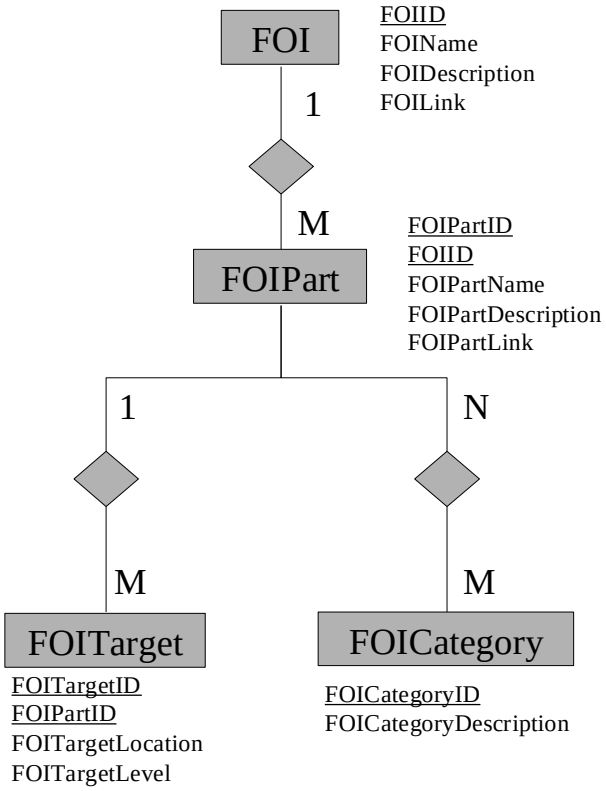


Figure 2.4: Entity Relationship Diagram for FOI Database

The allocation of category classes at FOI Part level allows for aspects of an FOI to be excluded or included in the results, depending on the user's preferences. Each FOI Part could belong to more than one category class.

2.5.2 Line of Sight Metrics

When exploring the urban environment a pedestrian's view is filled with features competing for attention. Some distant FOIs may be clearly visible, yet close items may be partially obscured. If these qualities are to be modelled then a set of corresponding metrics are required. These include a metric for proximity to the object, a metric for the amount of a feature that is visible, the field of view occupied, and to indicate if the FOI is on the skyline.

The approach used here was to consider an LOS from the observer to each of the FOI targets in turn, recording the vertical visible extent, and the location of close and distant horizons. These

results are then combined to form a number of metrics for each FOI part. The intention is that future LBS devices could use these values to deliver information to a user in a meaningful order, filtering out details not relevant to the current context.

2.5.2.1 Close Horizon

The 'Close Horizon' is calculated by locating the feature directly in front of the target which creates the steepest viewing angle between observer and target, not including the FOI itself. The elevation of this object, known from the DSM, is used to calculate the intercept of a line of sight with the target, and deduce how much of the target is visible (Figure 2.5 - *TH1)*. The obscured area is also recorded (Figure 2.5 - *TH2).*

2.5.2.2 Distant Horizon

By extending the LOS ray beyond the target it is possible to discover if the target makes the skyline, or is overshadowed by a more distant object. The search continues until either it intercepts the DSM, or reaches the same elevation as the maximum elevation in the DSM dataset. If the ray intercepts the DSM then the 'Distant Horizon' location is recorded along with the elevation value at that point, and both distance behind target and extent of the object showing *(*Figure 2.5 - *HD1)* are calculated. If no feature is found the target is designated as being on the skyline.

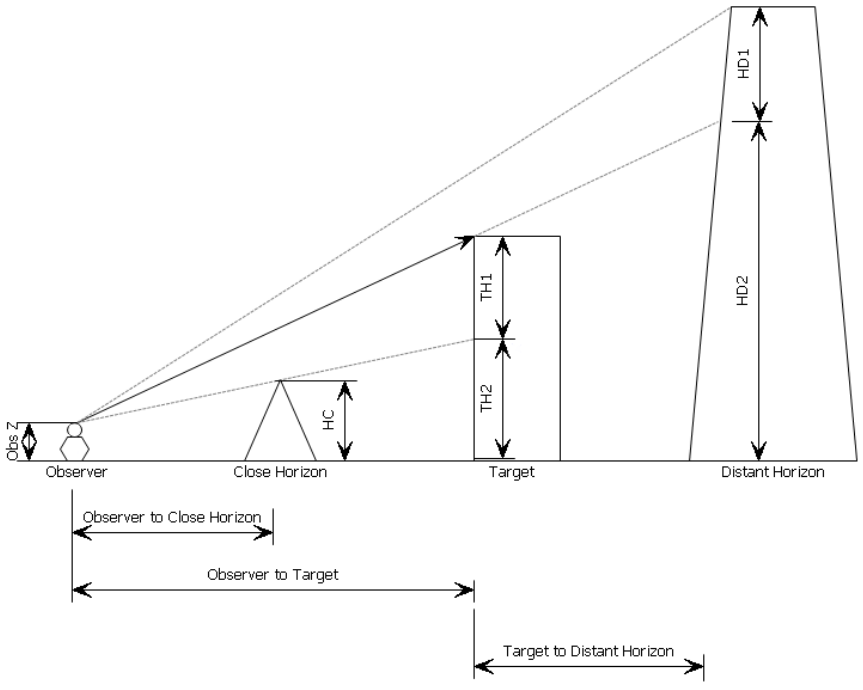


Figure 2.5: Line of Sight Details; the Close Horizon is used to calculate how much of the target is visible; the Distant Horizon information indicates whether the target is on the skyline or overshadowed by a taller object

The values stored from each of the LOS calculations between the observer and each target point are summarised in *Table 2-1*.

Table 2-1: Line of Sight Target Return Values

Criteria	Data Type	Details
Visible	Boolean (True or False)	Whether a target point can be seen or not from current location
Target Location	Point (x,y)	The location of the target point
Target Elevation *(TH1+TH2)*	Metres (vertically)	The height of the target point
Distance from Observer *(Distance O-T)*	Metres (horizontally)	The distance from observer to target point
Close Horizon Location	Point (x,y)	The location of the tallest object in the line of sight from the observer to the target
Close Horizon Height *(HC)*	Metres (vertically)	The height of the tallest object between observer and target
Close Horizon Intercept With Target *(TH1)*	Metres (vertically)	The amount of target which shows above the tallest near object
On Skyline	Boolean (True or False)	Whether the target has a taller visible object behind it, or sky
Distant Horizon Location	Point (x,y)	The location of the intercept with a taller object behind the target (if any)
Distant Horizon Distance Behind Target *(Distance T-DH)*	Metres (horizontally)	The distance from the target to the horizon
Distant Horizon Elevation *(HD1+HD2)*	Metres (vertically)	The elevation of the item on the horizon visible behind the target
Distant Horizon Intercept *(HD1)*	Metres (vertically)	The amount of the horizon that is visible above the target
Elevation Showing as Ratio Of Distance to Observer Ratio *(TH1 /Distance O-T)*	Ratio	Considering any near horizons, calculate the ratio of visible vertical extent divided by distance to observer

2.5.3 Visibility at Feature Level

For a LBS to make use of the target visibility attributes a number of summaries at the FOI Part level are required. These summaries Table 2-2 are intended to provide a LBS with the facility to filter spatial databases, and to sort results according to various quantitative measures. As the LBS is able to determine automatically which aspects of a feature are visible, it can customize the information delivered. For example it is able to report details of the highly visible spire, but not mention the currently obscured entrance lobby.

Table 2-2: Visibility Metrics at FOI Part level

Criteria	Data Type	Description
Visible	Boolean (True or False)	Shows if part of a feature can be seen
Average Distance of Visible Targets	Metres (horizontally)	The average distance to only those targets visible to the observer
Maximum Horizontal Field of View	Degrees (horizontally)	The field of view between the widest visible targets on an FOI
Number of Visible Targets	Integer	A count of the number of visible targets
Percentage of Targets Visible	Decimal	Count of targets visible divided by all targets on FOI Part
Total Face Area Visible	Square Metres	The combined total area of feature frontage visible, when considering close horizon
Total Face Area Blocked From View	Square Metres	The area on the frontage which is in the shadow of near blocking objects
Percentage of Targets On Skyline	Ratio	Count of visible targets on skyline, divided by all visible targets
Average Visible Target Height	Metres (vertically)	Average elevation calculated from all visible targets
Minimum Visible Target Height	Metres (vertically)	Lowest visible target elevation
Maximum Visible Target Height	Metres (vertically)	Highest visible target elevation

2.5.3.1 Field of View

Most of these metrics are self-explanatory, however 'Field of View' and 'Total Face Area' require clarification. The 'Field of View' is calculated between the most extreme targets visible, which make the widest angle. If an object obscures a side of the FOI frontage then the FOV angle will decrease. However if an object blocks a portion in the middle of the FOI with the outside targets still visible, the FOV will return the widest viewing angle calculated from the outside points, ignoring the obscured portion of the FOI frontage.

2.5.3.2 Total Face Area – Visible / Blocked

The 'Total Face Area' visible is calculated by summing the area visible under each target after considering the area obscured by obstacles between the observer and target, as illustrated in Figure 2.6. The difference between DTM and DSM establishes FOI height, removing topography from the resulting area. The Total Face Area Blocked reports the area which is shadowed by near objects.

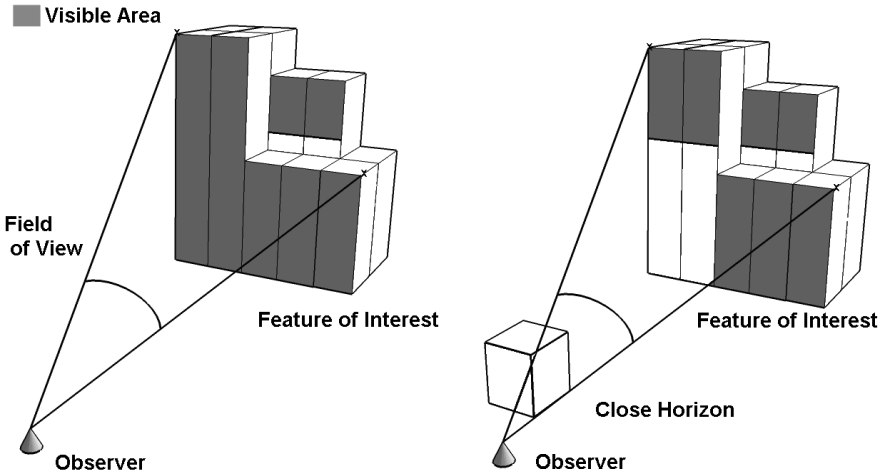


Figure 2.6. Total Face Area Visible for a Feature of Interest

A number of derived metrics may also be useful such as the ratio of distant horizon height showing (HD1) over distance behind target, to give an indication of dominance of any distant features. Also the total visible face area divided by the average distance to the targets would give an indication of the presence of an FOI from the user's viewpoint.

2.6 Implementation and Evaluation

An initial demonstration of the method was carried out in the city of Christchurch, New Zealand. Christchurch Cathedral was selected as an FOI, and divided into 3 component parts as outlined in Section 2.5.1. A number of observation points were selected to give different views of the Cathedral, photographs were taken at each site, and the GPS locations passed to the algorithm to return the visibility metrics. Figure 2.7 shows the location of the test sites, along with corresponding photographs.

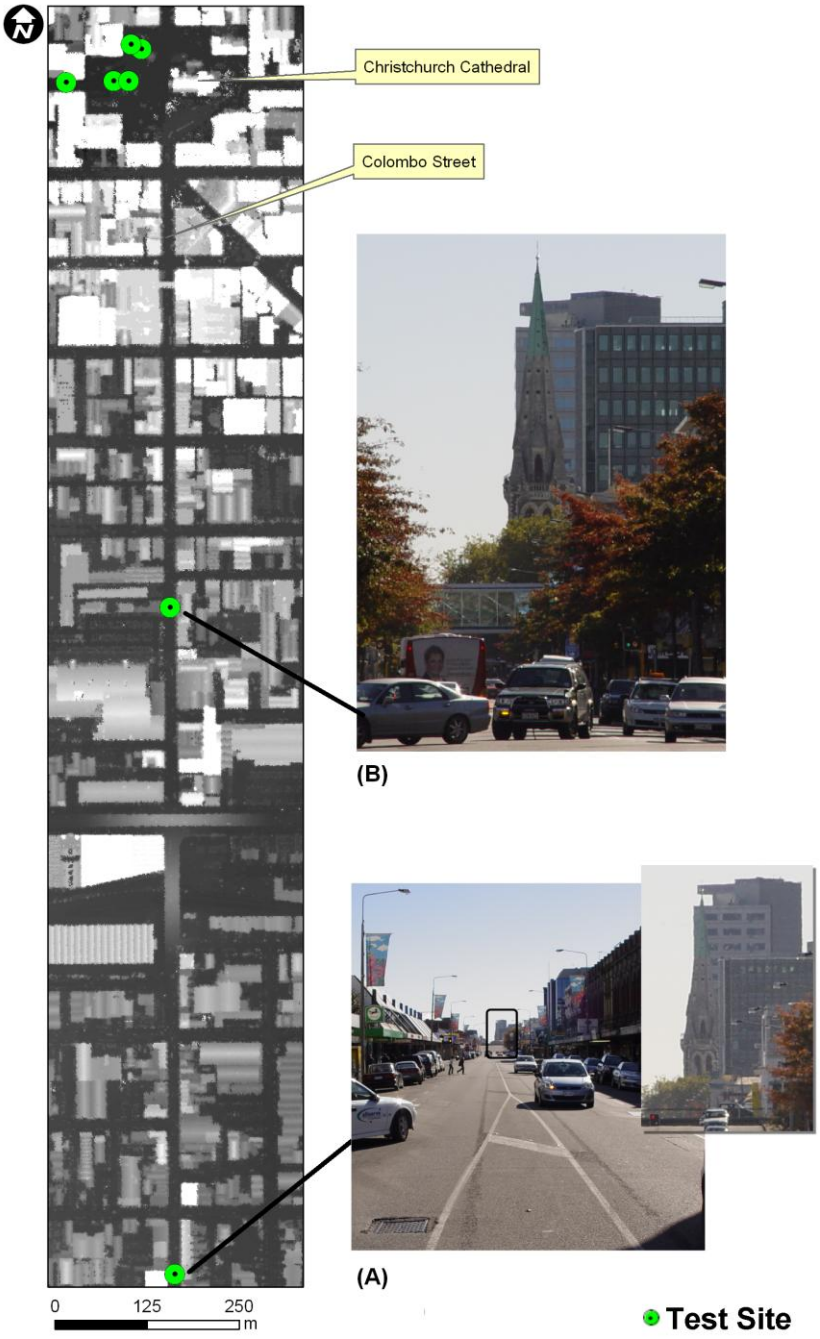


Figure 2.7: Map of Test Sites in Christchurch, New Zealand (LiDAR data sourced from Christchurch City Council)

Table 2-3 shows the results from running the visibility calculations. The third part of the FOI was not visible from either location A or B, and has been removed from the table.

Table 2-3 Results from Sites A and B

Location		A		B	
		Part 1	Part 2	Part 1	Part 2
Visible Targets (%)		18.2	0.8	17.7	0.6
Percentage Elevation Showing of the Visible Targets (%)		13.4	8.2	12.9	7.9
Total Face Area Visible (sq m)		202.5	33.4	191.5	28.0
Total Face Area Blocked From View by Near Horizon (sq m)		376.1	1325.9	318.7	1316.5
Average Distance of All Visible Targets (m)		1636.5	1620.5	721.6	706.4
Percentage of Targets on Skyline (%)		3.0	0	2.5	0
Visible Target Height (m)	Average:	41.3	25.6	41.9	26.6
	Maximum:	59.1	29.0	59.1	29.0
	Minimum:	20.1	22.0	26.1	22.0
Maximum Horizontal Field of View for Entire FOI (degrees)		0.25	0.11	0.56	0.24

The majority of the results are as expected with the more distant site showing the FOI to occupy a narrower field of view, that the spire (FOI Part 1) is visible, and none of main building (FOI Part 2) makes the skyline due to the tall surrounding buildings.

However the results show that the percentage elevation of visible targets (13.4% at A, 12.9% at B), percentage of targets on the skyline (3% at A, 2.5% at B), and total face area visible ($202m^2$ at A, $191m^2$ at B) for the spire (FOI Part 1) go against the expected trend and are slightly greater at Location A than B. It is also noticeable that a greater extent of the spire is visible at location A (20.1m to 59.1m at A, and 26.1m to 59.1m at B).

In fact, although counter-intuitive, these values match the real world experience as seen in the photographs taken from these sites. At Location A the vertical extent of the spire visible is greater with the majority of the left side making the skyline, as annotated in Figure 2.8. At Location B the distant skyscraper blocks the sky behind the spire, and trees in the foreground obscure the lower aspects of the spire. This is in agreement with the output values from the algorithm.

Location A **Location B**

Figure 2.8. Comparison of Cathedral Spire from Locations A and B

2.6.1 Additional Analysis

Five further test locations (Figure 2.9) were selected and the metric results calculated. Values for total visible area frontage along a transect line passing from Location C, through D, and E are shown in Figure 2.10.

A statue and several trees block the view of the Cathedral along this approach, as reflected by the metrics.

Considering the percentage elevation of targets visible, at Location C 39.9% of the spire is visible, at Location D the main body of the Cathedral receives a score of 17.2%, whilst the spire receives a value of 11.8%, indicating the spire is obscured more than the main body. At Location E the main body received a score of 20.4%, while the spire receives 34.8% indicating the prominence of the spire once more.

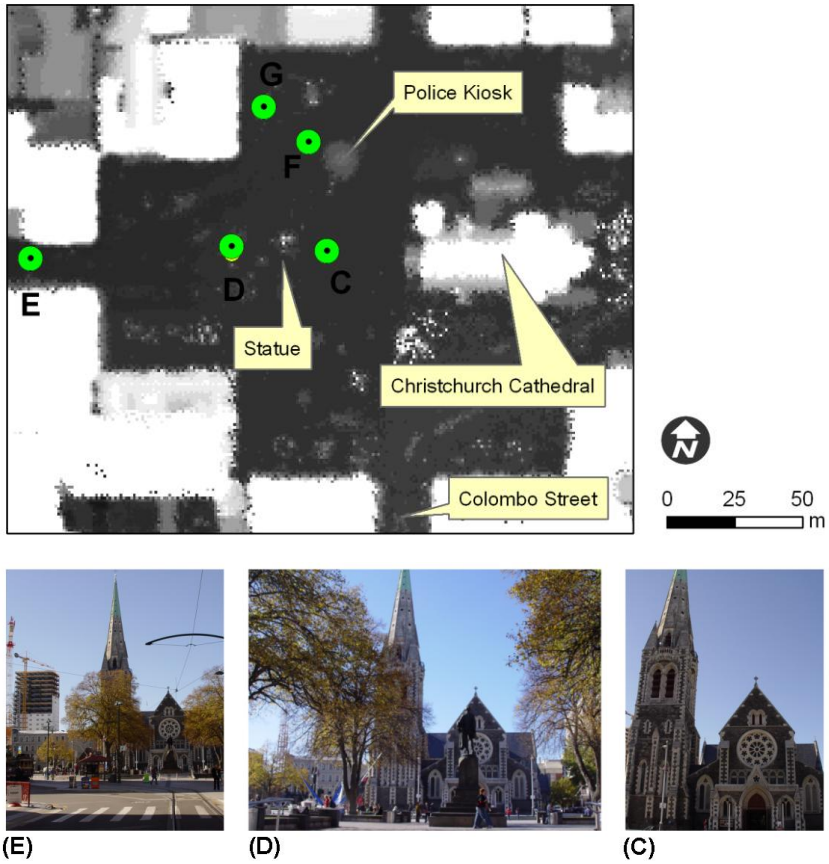


Figure 2.9. Cathedral Square Test Sites (LiDAR data sourced from Christchurch City Council)

The main difference in the view between locations F and G is the proximity of a Police hut (Figure 2.11). At Location F, the Café (FOI Part 3) is not visible, whilst at Location G it scores 6.5% (percentage elevation of visible targets).

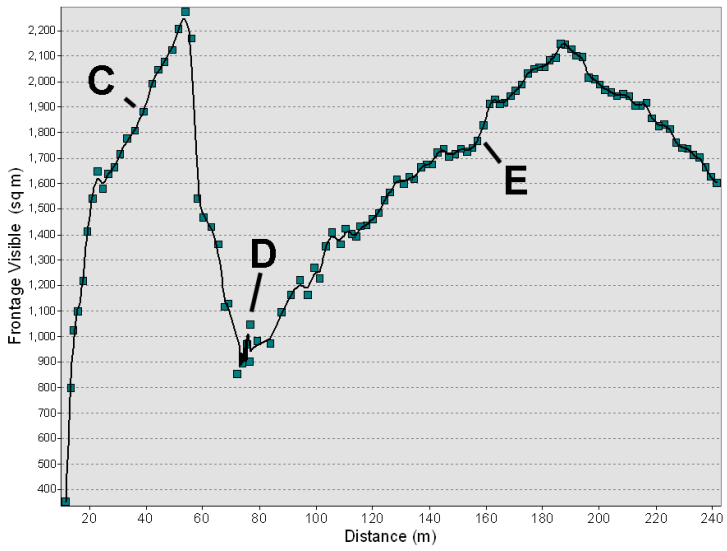


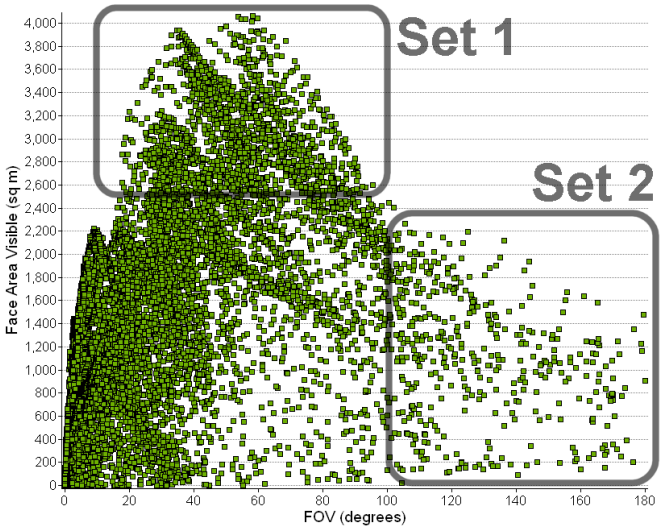
Figure 2.10. Frontage Area Visible against Distance from Feature of Interest

The total face area of the spire at Location F is 615.1m^2, whilst at Location G it receives a value of 818.3m^2. These are in agreement with the photographic evidence.

Location F **Location G**

Figure 2.11. Site F and G photographs

There is an interesting relationship between the field of view, and total area visible. Figure 2.12 illustrates this by considering 26,753 locations around Cathedral Square spaced 1 metre apart.



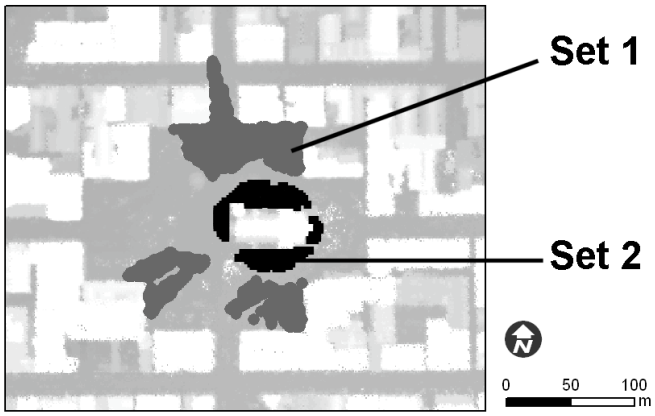


Figure 2.12. Field of View and Face Area Visible (LiDAR data sourced from Christchurch City Council)

Close to a feature the field of view metric scores high, and the total frontage face area visible is low. As can be seen from the graph the points which contribute to the peak of face area value points (Set 1) are located away from the Cathedral on the edge of the square, whilst the highest FOV values (Set 2) are near the Cathedral. From a viewing experience the face area values may be considered the most appropriate metric to reflect 'how much' of an FOI can be seen, and should be considered with distance to quantify the presence of an FOI.

2.6.2 Mapping the Visibility Metrics

The values from the LOS implementation may be mapped to indicate how a user's experience of an FOI would vary across space (Figure 2.13). In this example the area in front of the Cathedral between test locations C and D (Figure 2.9) enjoys the highest visible percentage of targets, essentially the clearest view.

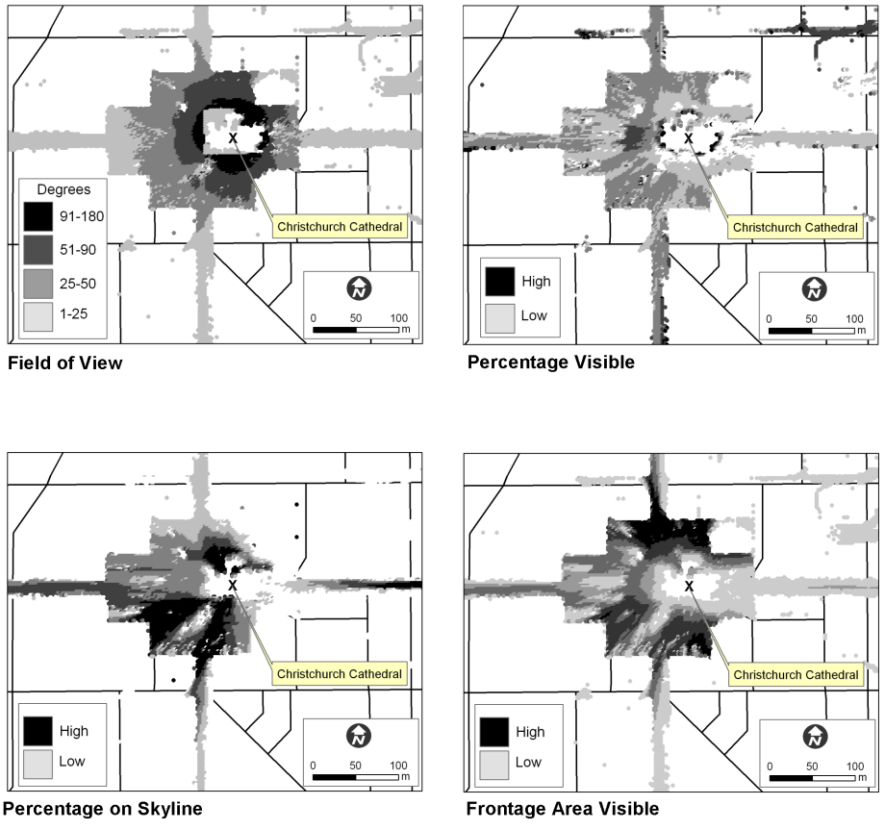


Figure 2.13. Maps of Key Metrics

Calculating the gradient from any of these datasets will show the magnitude and direction of change of the metric, and may be used to indicate in which direction a user should move to see more, or less, of the FOI.

2.6.3 Using the Visibility Metrics

These metrics combined with other GIS datasets can provide an LBS with the ability to be context-aware when delivering information to a user. When searching spatial databases the visibility metrics may be used to rank the results showing the most visible items first. In way-finding, the visible area values may be used to lead the user towards good viewpoints for nominated landmarks. Any information delivered to the user would need to be supported by additional datasets, such that attributes including the FOI's name, address, usage, and history could be conveyed to the pedestrian.

Although not implemented at this stage, a fuzzy logic layer will be introduced in the next phase of the research so that a natural language engine may select the most appropriate English terminology to describe the scene. This will allow the LBS to take on the role of a virtual city guide. The class limits will be set after conducting user trials to evaluate perceived object visibility against the metric results.

To ensure the LBS remains responsive to the user's movements, the metrics must be available in real time. This can be achieved in a number of ways. One method is to pre-cache the results for an area of interest, such that the mobile client requires only a simple lookup of the corresponding FOI visibility summaries.

An alternative approach, which will be used for the next phase of this research, is to implement a client-server architecture whereby a mobile client sends the user's location information across a network to the server which returns the visibility results for the surrounding region. Pre-caching on the server side is also possible such that the most commonly visited areas can be held in memory for improved response times.

2.7 Conclusion and Future Work

This research has shown that a line of sight algorithm may be used to supply a number of useful contextual visibility metrics from a LiDAR dataset in an urban environment. These results form part of a function of visibility which can be used to prioritise information on features of interest from the current observation location.

We have shown that it is possible to incorporate a wide range of visual statistics into reports of object visibility, with consideration of the surrounding cityscape, and the importance of close and distant horizons.

There are a number of areas for further research in this field. The algorithm considers only the physical aspects of visibility which can be calculated from a DSM. It currently neglects to consider the time of day, or weather in the visibility calculations.

It would also be beneficial to measure the texture, material, colour and contrast differences between the target and any distant horizon objects, such that a metric may be established to

indicate how easily an FOI may be resolved from its background. Aerial imagery might offer a partial solution suitable for identifying vegetation backgrounds, but roof top colours will not assist in contrast and colour information from the user's viewpoint, so georeferenced street level photography would be preferable.

Partial visibility through vegetation (Llobera 2007a) could be explored such that lines of sight are able to pass through, and under, canopy layers. The algorithm should be extended to accommodate seasonal tree canopy and vegetation density variation. It would also be worthwhile to examine raw LiDAR returns to produce detailed surfaces for the canopy layer.

Currently a scan in the vertical axis over the FOI returns information on the distant horizon. This could also be applied across the horizontal axis so that the surroundings may be used as context to determine if an FOI profile is significantly different from its neighbours (e.g. skyscraper amongst low rise buildings).

Referencing a number of FOIs in the metrics would allow a LBS to use relative descriptions of space, such as "the Chalice monument is visible to the right of the Cathedral". Accessing models from the Google 3D Warehouse may also be useful such that pictorially the shape and texture of FOIs could be displayed, allowing the user to more easily identify the target FOI from surrounding buildings.

The ultimate goal of this research would be a single visibility function which considers weighted metrics based on visibility, architectural interest, building form, and social factors such as building use or related interests to the user, that would allow an LBS to rank in order from a scene those items which a viewer would consider most interesting. The values would then be passed through a fuzzy logic layer and a natural language engine to generate appropriate sentences to convey to the user descriptions of their surroundings, such that the LBS acts as a virtual city guide.

2.8 Acknowledgements

The authors would like to thank the Christchurch City Council for use of their LiDAR dataset, and ZNO for the 3D model of Christchurch Cathedral. The research would not have been possible without funding from the University of Canterbury, New Zealand. We would also like to thank Dr William Mackaness for comments on an early draft of the chapter.

Chapter 3: Advancing Visibility Modelling Algorithms for Urban Environments

3.1 Summary

This chapter presents a number of methods for calculating the visibility of landmark features in the urban context. While current visibility modelling techniques establish which regions may be viewed from a location, it is sometimes necessary to quantify how much of a designated feature is visible from its surroundings. This is particularly relevant in the field of Location Based Services, where information is currently filtered using proximity and syntax matching, but could include more advanced egocentric and contextual filtering capabilities if feature visibility modelling was available. This research presents a method to establish the visible extents of landmarks in an urban environment, through the development of a number of visual metrics. These metrics are tested in both synthetic and real world trials.

3.2 Introduction

Standing in Trafalgar Square (London) it is possible to see Nelson's column, The National Gallery, and many statues and fountains. The space is largely defined by the visual field; which icons can be seen, their relative locations and how clearly they are in view. These landmarks are prominent features because of their visual, semantic or architectural attraction (Raubal and Winter 2002), and form useful reference points (Millonig and Schechtner 2007) which assist in building a mental model of the region useful for navigation and exploration (Fisher-Gewirtzman and Wagner 2003). Currently Location Based Services (LBS) under-use landmark visibility modelling (May *et al.* 2005), and rely on more simplistic proximity filters. However by introducing visibility modelling a LBS would be able to more closely match the user's experience and determine the impact of landmarks within the visual field. This chapter presents a novel method for quantifying landmark visibility with the eventual aim of providing the foundation for natural interaction with the user that aligns with their mental models, and matches their visual experiences.

The advent of high resolution Light Detection and Ranging (LiDAR) sourced terrain models, which capture building form and topography, has extended the scope of visibility modelling applications. LBS is one area able to benefit, these applications determine relevant information by integrating the user's location with other datasets (Spiekermann 2004). This definition includes mobile phone applications able to assist the user in locating nearby entities (e.g. banks, restaurants, friends), in-car navigational devices, historical city guides, location based games,

virtual fitness training partners, and mobile commerce (e.g. customised local advertising or targeted discount coupons).

Typically filtering is performed using syntax matching and feature proximity as search inputs, however more relevant results may be found when the user's context is also considered, such as the time of day, or current weather conditions (Yu *et al.* 2005). The incorporation of visibility analysis allows the LBS application to model the user's environment more closely and determine which surrounding features are in view, and therefore supply more meaningful assistance. For example landmarks have been recognised as useful navigational points (Millonig and Schechtner 2007), and by modelling their visibility an LBS may announce them as they come into view (Elias and Brenner 2004), leading to more natural navigation narratives.

While most Geographic Information Systems (GIS) provide facilities to calculate the extent of visible space from a given point (De Smith *et al.* 2007), they do not provide the functionality to model the visibility of a landmark across space. This is addressed in the concept of visual exposure presented by Llobera (2003) by modelling the field of view occupied by a landmark from any surrounding location. This concept was further developed through the introduction of a number of additional metrics to quantify the visibility of landmark buildings within the urban environment (Bartie *et al.* 2008). This chapter continues that work through synthetic and real world trials, resulting in improvements to five of the visibility metrics, each described and demonstrated in the urban context. These metrics are designed with the intent that they may be used in future LBSs, such that consideration is given to contextual information regarding the user's situation, their viewpoint and the relevance of surrounding landmarks. In so doing announcements may be made in order of landmark visual prominence, or by referring to particular visual attributes (e.g. 'the *large* tower on the *skyline*'). The metrics may be used to establish, from standard GIS datasets, how much of a feature is visible given in terms of the surface area exposed, or how large an object appears when considering the viewing distance. Such metrics would permit a LBS to establish the visual dominance of a feature, so for example features receive attention appropriate to their perceived size. Calculations of landmark clarity are also possible by considering the location of visual blockades, enabling the use of terminology such as "very clear" or "mostly hidden" to be introduced into narratives. Also it is possible to find the highest and lowest visible points on a feature, and to calculate how much appears on the skyline.

Incorporating visibility analysis means LBSs are able to more closely model the surrounding urban form and user's context so that results may be sorted according to their visual relevance and features which are out of sight may be disregarded. To do this, visibility modelling at feature level (i.e. landmark buildings) needs to be developed further. This chapter presents a modified LOS algorithm (Section 3.4) able to establish a wide range of visibility metrics for nominated features (Section 3.5), with the ability to map which parts of a feature are visible. The algorithm

was tested with a number of synthetic trials (Section 3.6) before exploring its use in the real world (Section 3.7).

3.3 Background

Previous research on visibility has been divided into urban and rural disciplines. Studies in the urban landscape have tended to be based on isovists (Tandy 1967), using in particular Benedikt's (1979) interpretation and definitions. Essentially 2D isovists describe the space which is visible from a vantage point as a closed polygon. Consideration is given to the form of the built environment through the use of architectural plans which denote the building footprint and position, however building height is ignored. The topography of the land surface is disregarded, as is the continuation of the lines of sight beyond the first intersection with a building footprint. Therefore isovists depict lines which when traversed from the vantage point offer a continuous view of the target, and disregard more distant features. More recently there have been developments in 3D Isovist theory (Morello and Ratti 2009), which access a Digital Elevation Model (DEM) to consider building form and topography. These may be used to describe the space around an observer, using metrics such as openness, and maximum viewing distance.

In the rural context, viewsheds have been used to show regions which are visible from a vantage point (Tandy 1967, Lynch 1976). These are created by calculating which cells of a DEM are visible from an observation point, using a line of sight (LOS) algorithm. They are also used in military and tourist applications for finding the most hidden route, or those with the best views (Lee and Stucky 1998). More complex forms include complete intervisibility databases (Caldwell *et al.* 2003) able to report how often a region is visible from the surrounding space, and visibility graphs (O'Sullivan and Turner 2001, Turner *et al.* 2001) which describe the pattern of mutually visible regions in a landscape.

However to model a feature from its surroundings, as required by an LBS, an alternative model based on calculating 'visual exposure' (Llobera 2003) is used. This estimates how much of a feature can be viewed from the surrounding space, enabling the creation of surfaces to show in which direction an observer would need to move to view the target more, or less, clearly. In Llobera's work visual exposure is a measure of the horizontal and vertical fields of view occupied by an object from the surrounding locations. This technique can be used to find visual corridors, or visual ridges, and forms a useful basis for considering feature visibility in the context of LBS.

The model presented in this chapter builds on the visual exposure model, establishing a range of visual metrics able to describe the visibility of a nominated Feature of Interest (FOI). This differs from the 3D Isovist approach which quantifies the space around the observer, as here the attention is on how much of a target feature is visible. The visual exposure model is developed

through extending the basic LOS algorithm to provide more information on how much, and which parts, of a feature are visible. Groups of cells within a defined boundary are treated as single objects, enabling the visible exposure of entire features to be modelled. The calculations are based on a Digital Surface Model (DSM) generated from a LiDAR dataset, able to capture high resolution height values including topography, vegetation and building profiles. These have been shown to be suitable for such applications in the urban environment in previous visibility studies (Palmer and Shan 2002, Bartie and Mackaness 2006). LiDAR models were also used to good effect in a navigation application by Elias and Brenner (2004) to assess the visibility of landmarks. The approach they adopted was to render perspective views of scenes, with each landmark drawn in a unique single colour so that simple image processing techniques could be used to quantify the visibility of each building. The technique benefits from being able to use the graphics card's processing power, but is limited by the range of metrics available and integration with existing spatial algorithms. Also the results are generated at a feature level and it is not possible to drill down into the raw data to determine which individual parts of a feature are visible, for example if the entrance to a building is visible. It was for these reasons a new model was developed able to report the visual exposure for FOIs.

3.4 Visibility Modelling in Urban Regions

When considering the visibility between an observer and a target, a LOS is calculated which reports a Boolean visibility result indicating whether the target cell is visible or hidden. By defining the feature boundary a collection of cells can be considered as a single object, such that a visibility score may be calculated for the entire feature. Each landmark is considered as a Feature of Interest (FOI), and not the more common Point of Interest, as the implementation used here summarises the visibility properties for a collection of cells rather than a single point. A simple example of this is given in Figure 3.1, where each cell is allocated a value of 1 if that cell is visible from the observation point. At the feature level this may be summarized by adding up the total number of cells visible and dividing by the total number of cells within the defined boundary, to provide a visibility percentage for the feature (Caldwell *et al.* 2003).

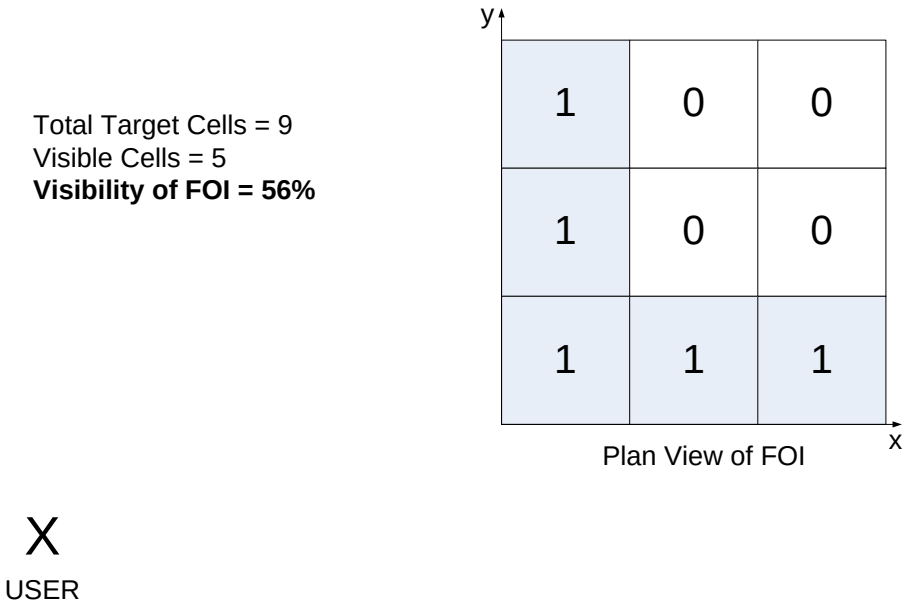


Figure 3.1: Calculating Boolean Visibility for Each Cell within a Feature of Interest

Although topography and building elevation are considered during the LOS calculation, the vertical visible extent for each cell is not calculated. An adaption to the LOS algorithm (Bartie *et al.* 2008) makes it possible to consider the most prominent feature in front of each target cell, that is the feature which creates the steepest viewing angle between user and target. From this the interception height of the LOS with the target may be calculated, and therefore the vertical extent of the target cell showing above this intercept indicated as TH1 in Figure 3.2.

Additionally it is possible to determine those target cells on the skyline by extending the LOS beyond the target, until either it intercepts the DSM or reaches the same elevation as the maximum found in the DSM dataset. If the ray intercepts the DSM then the 'Distant Horizon' location is recorded along with the elevation value at that point, and both distance behind target and the visible vertical extent of the horizon point are calculated. If no 'Distant Horizon' is found the target is designated as being on the skyline. Repeating this process for all of the targets on the FOI gives an indication of how much of the feature is on the skyline from a viewpoint.

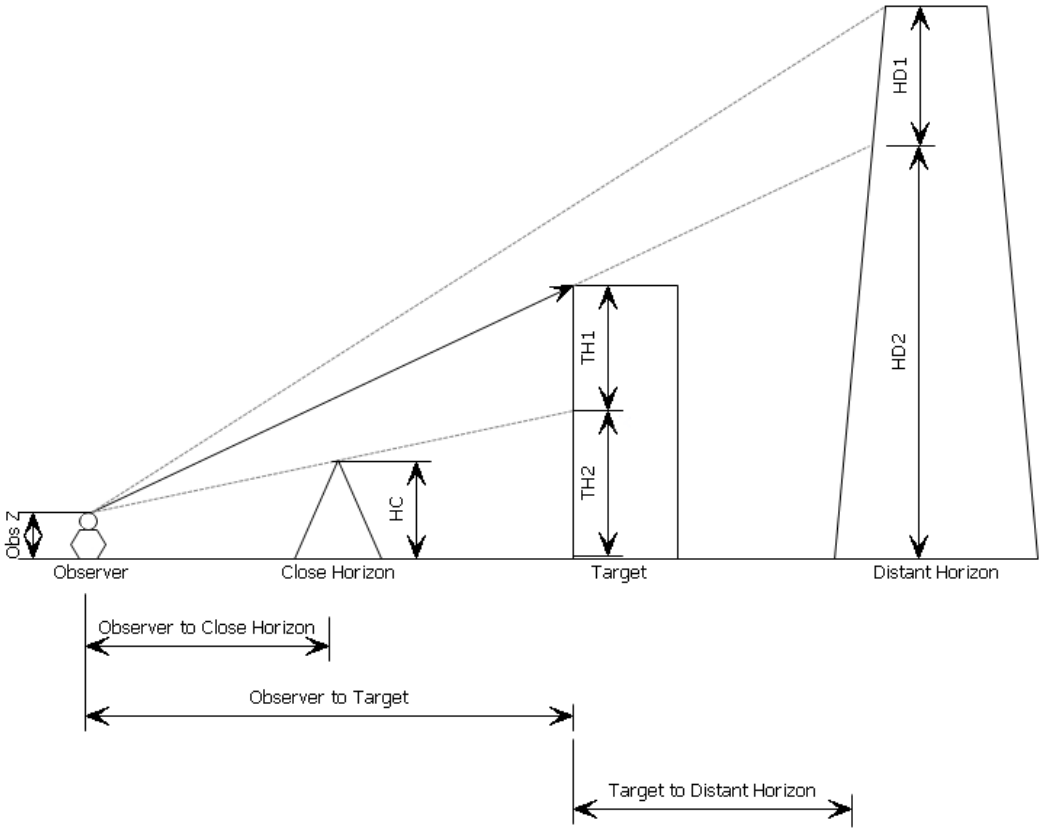


Figure 3.2: Modified Line of Sight Algorithm (Bartie et al. 2008)

HC is the recorded height of the close horizon, TH1 is the height of the visible portion of the target while TH2 is that hidden from view by the Close Horizon. HD1 is the visible extent of any overshadowing object behind the target, while HD2 is the extent obscured.

Figure 3.3 shows how the modified LOS algorithm more closely models landmark visibility by including the feature's profile. Each cell shows the DSM elevation, and the amount visible calculated from the LOS interception heights.

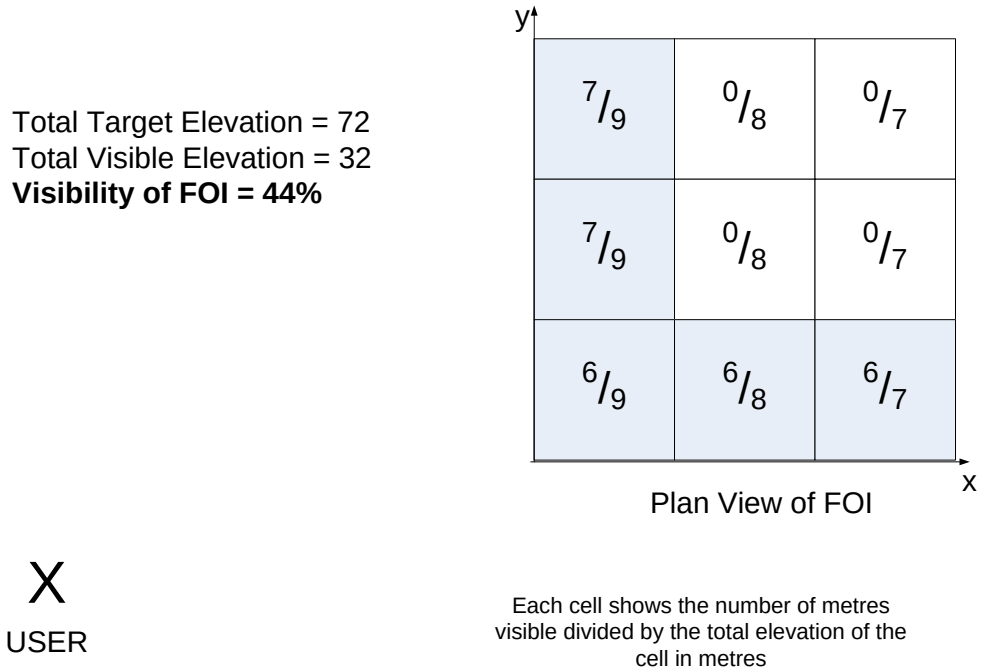


Figure 3.3: Visibility of a Feature of Interest with Consideration for Vertical Extent

One further step is required to establish the visible percentage of a FOI, by separating the elevation of the feature from the topography (Figure 3.4). This calculation can be performed by subtracting the bare earth elevation value, available from a Digital Terrain Model (DTM) sourced from LiDAR data. The result is an algorithm able to report target visibility, with consideration to building form, able to denote skyline targets, and report the percentage of a FOI which is currently visible.

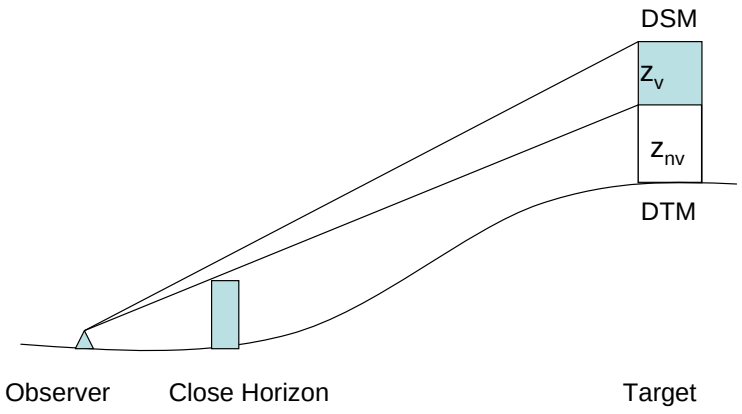


Figure 3.4: Consideration of Topography when calculating Object Visibility.
Z_v denotes visible elevation, while Z_{nv} denotes the extent hidden from view.

3.5 Extensions to Existing Visibility Models

Each FOI consists of a set of cells within a defined boundary; therefore summaries can be calculated to reflect the visual properties for the entire FOI from a given viewpoint. In these trials one metre resolution surface models are used, with every cell within the FOI boundary considered as a target for the LOS calculations. The summaries indicate the total field of view occupied by the FOI, façade area visible, the area hidden from view, and how much of the FOI is on the skyline. It is also possible to report the maximum and minimum elevations visible, and thus deduce if the observer can see only the top of a FOI behind other buildings. The next section outlines the methods for calculating the summaries.

3.5.1 Calculating the Field of View Metric

The 'Field of View' (FOV) is calculated as the horizontal angle between the most extreme FOI targets visible, which make the widest angle from a given viewpoint. It is a measure of object size and viewing distance, but does not indicate the extent to which the FOI interior may be obscured. Therefore a façade area is also calculated by summing the visible extents under each target taking into account any area obscured by obstacles between the observer and target. Figure 3.5 shows how the FOV remains unchanged while the visible area decreases when a close object is added to the scenario.

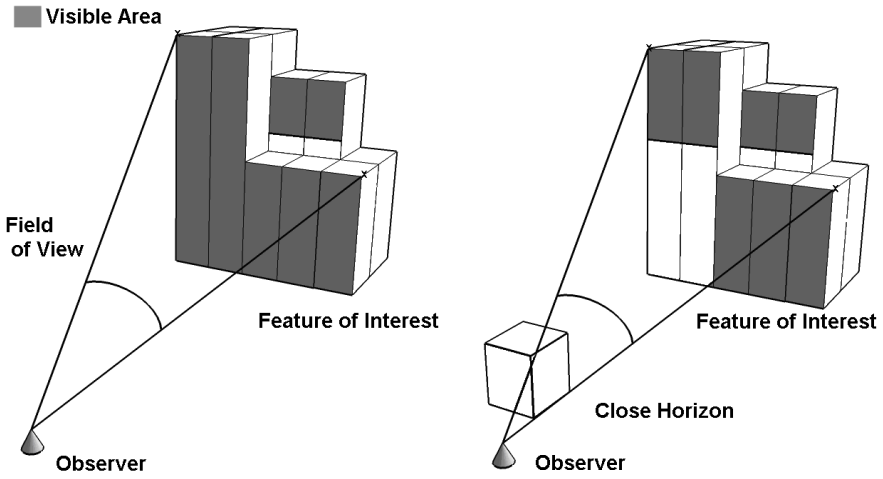


Figure 3.5: Field of View and Total Façade Area Visible for a Feature of Interest

3.5.2 Calculating the Visible Area Metric

The area calculations consider the total FOI façade area which can be viewed from a given location, taking into account the regions obscured by closer features but irrespective of viewing distance. To better establish the visual dominance of each FOI the effects of distance and viewing angle must be considered to generate a "perceived visible area" metric.

To clarify the difference in these definitions consider an A4 sheet of paper, which has a constant façade area yet when viewed from different angles or distances has a varying perceived area (Figure 3.6). In the model each FOI target is considered as a vertical column with a width of one metre, and a vertical extent defined by the object height and LOS intercept. The relative angle between the user and the FOI target column determines how wide that column appears to the observer. This only provides part of the adjustment for viewing angle though, as the DSM is stored as a raster dataset where each cell is square and aligned directly north-south. Therefore any FOIs which are not orientated with the grid direction are constructed from diagonal lines represented by patterns of raster cells. As an example, if a 14m long wall was digitised at 45 degrees from the grid orientation, it would occupy 10 raster cells of 1 metre each. Any calculations which ignore the original feature angle, and calculate length by multiplying the cell size by the number of cells in the mapped feature would underreport length (i.e. 10m instead of 14m). Therefore each FOI was processed using a 3 by 3 kernel to detect edge patterns and establish the exterior wall angles (Figure 3.7). Results were assigned to the central kernel cell, and interior cells were assigned an angle based on the nearest calculated edge cell.

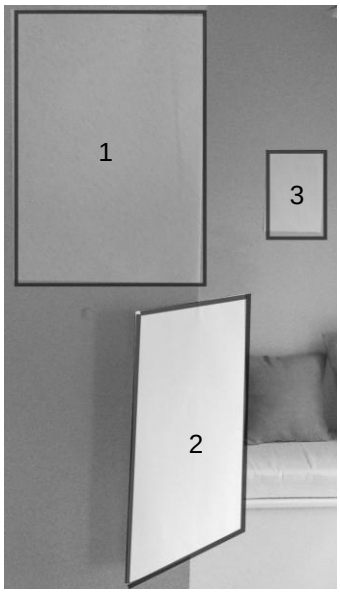


Figure 3.6: Viewing Angle and its Effect on Perceived Area

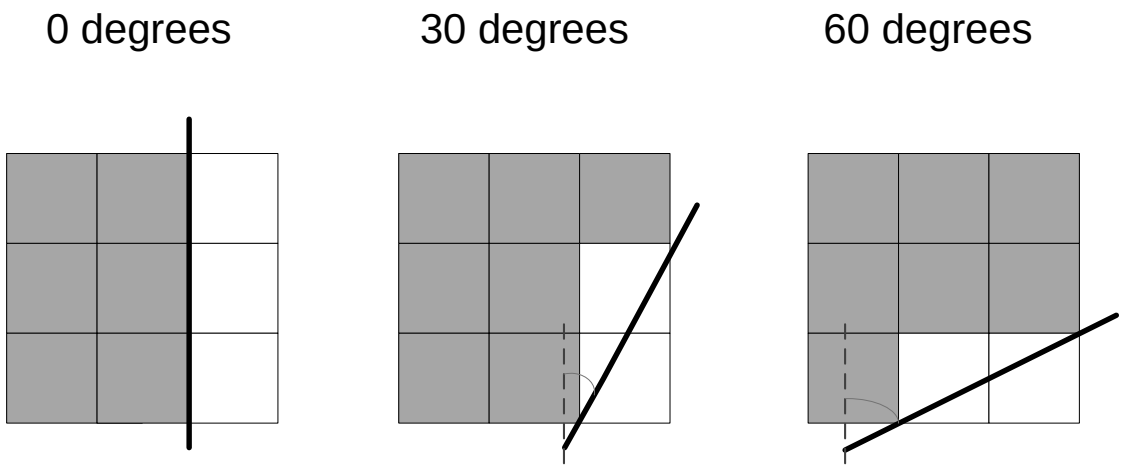


Figure 3.7: Examples of Detecting Edge Angles from Raster Dataset using a 3 by 3 Kernel

Using both the viewing angle from the observer to the FOI, and the information on the angle of each FOI façade the overall perceived area may be calculated. This is repeated for all visible targets resulting in the perceived surface area appropriate to the viewing angle, as given by

$$P = \sum_{t=1}^{n} A_t \mid Sin(V_t - W_t) \mid , \qquad \textit{(eq.3.1)}$$

where *t* is the target cell inside FOI boundary being considered, *A* is the area visible for the target (sq m), *V* is the angle in degrees between the observer and the FOI, *W* is the FOI façade angle (degrees), and *P* is total perceived area (sq m) for the FOI from the current location.

3.5.3 Introducing Distance Factors into Perceived Area Calculations

It can be argued that viewing distance may be excluded in calculating the perceived area of an object because of a characteristic known as size constancy (Boring 1964) which occurs when an observer considers an object's size to remain constant at a range of viewing distances. For example an observer might consider two cars on a street to be similar sizes, even though one is further away and occupying a smaller angle in the visual field. This is believed to be a result of the brain recognising the object in context and not just considering the space it fills on the retina. Applying this logic to the FOI visibility calculations would mean that the area visible may be considered irrespective of distance. However to be able to consider the dominance of a feature in comparison to others in view, a perceived area calculation which takes into account a distance factor is included in the form,

$$Perceived\ Area\ Including\ Distance = \frac{\text{Perceived Area}}{\text{Distance}^2}, \qquad \textit{(eq. 3.2)}$$

based on the premise that the size of an object in the retinal image is equal to the size of object divided by the distance to the object (Schlosberg 1950).

3.5.4 Calculating the Clearness Index for a FOI

A clearness index is calculated to establish if the user has an uninterrupted view of the FOI. It is defined as the total visible area expressed as a proportion of the total area which could be viewed if all other surface objects are removed from the scene. This is done by checking the location of the close horizon point (Figure 3.2) from each LOS against the FOI boundary to determine whether it is within the FOI itself or outside, as outlined in Figure 3.8. If it is within the FOI boundary then the target is considered to be permanently blocked from that viewpoint, due to the structure itself. However if the close horizon is found to be outside the FOI boundary it is disregarded and the full vertical extent visible for that cell is calculated. The outcome is that the total surface area of the building irrespective of surrounding features can be calculated from a viewpoint, such that the surface area visible may be expressed as a percentage of the total potential building surface area which would be visible without foreground objects. This may be used to express the extent to which aspects of the landmark are hidden from the viewer, or through carrying out clearness calculations from multiple locations to determine the clearest viewing locations for a landmark.

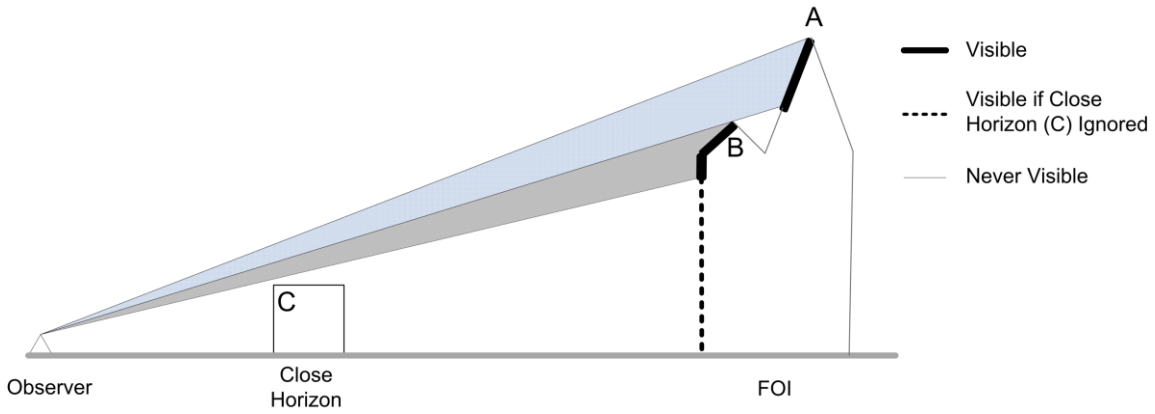


Figure 3.8: Calculating a Clearness Index

A LOS to Target A has a Close Horizon at Point B; A LOS to Target B will have a Close Horizon at C outside of the FOI boundary. Using the location of the Close Horizon points we can calculate which regions are always visible, those never visible, and those visible if the surrounding vegetation and buildings did not exist. These values are used to calculate the clearness index.

Hydrological functions may be applied to a 'clearness surface', constructed from multiple clearness calculations across a region, to establish the 'flow' direction. As hydrological models calculate the flow direction from high to low values the resultant map indicates the direction an

observer should move for a more limited view of the landmark. The direction to move for a clearer view is not the opposite of this. Figure 3.9 shows why this is the case, whereby moving from A to B would result in a decrease in the viewing clearness of the feature, however a better view from B would result from moving to C, and not back to A. Instead the direction an observer should move for a clearer view may be calculated by inverting the clearness surface before running the hydrological flow model, such that values of 0% indicate entire features are visible, while 99.99% represents highly hidden features. The resulting output specifies the direction to move for a clearer view of a feature.

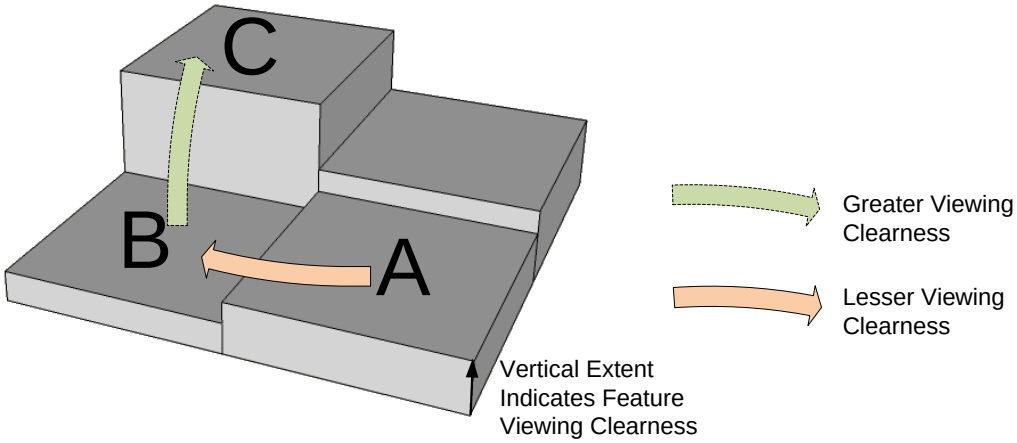


Figure 3.9: Feature Viewing Clearness – a move from A to B will result in a more limited view. However to obtain a clearer view from B a move towards C is recommended, not back to A.

3.5.5 Calculating how much of a FOI is on the Skyline

The skyline metric indicates if a target cell is overshadowed by taller more distant objects or on the skyline, using the LOS algorithm shown in Figure 3.2. Early trials indicated that in many cases the skyline summaries did not agree with the user's expectations. For example in Figure 3.10A the skyline algorithm reported that only 49% of the targets were on the skyline. The reason being that many visible targets (e.g. Figure 3.10B) were blocked from the skyline by more distant parts of the FOI structure. To quantify the human concept of an object being on the skyline it is necessary to establish a method for calculating the percentage of those targets visible which are overshadowed by objects outside of the FOI boundary, or not overshadowed at all. This gives an indication of how much of the FOI silhouette is on the skyline, as shown in Figure 3.10C. The model was updated to implement this concept by considering only points blocked by objects outside of the nominated FOI boundary as not on the skyline, ignoring visible points blocked by the FOI itself from the final skyline summary. The pseudo-code for this is shown in Section 3.5.6. The new algorithm was found to more closely reflect the skyline concept, with an updated skyline result of 98% for the building shown in Figure 3.10.

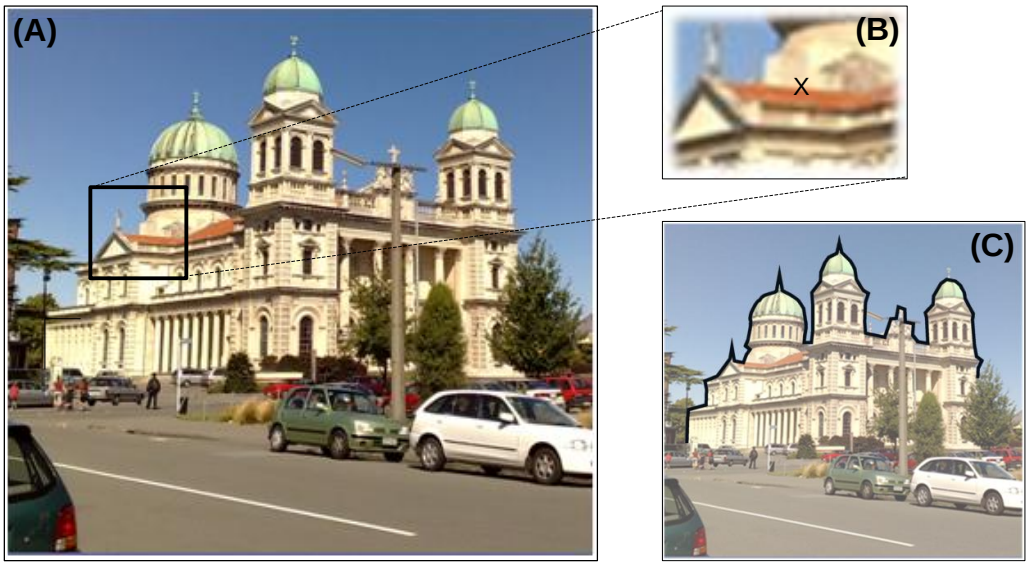


Figure 3.10: Real World Example of the Skyline Metric

(A)FOI from a given viewpoint (B) A target which is not on the skyline, as it is overshadowed by more distant parts of the FOI (C) Outline of FOI considered as the skyline value in updated implementation

3.5.6 Pseudo-code

The following pseudo-code outlines the computational steps taken when performing a LOS calculation for each target. The first method determines if the target column is visible, and if so the extent visible above the close horizon. It also calculates the clearness ratio by comparing results with and without foreground items. The second method reports if the target point is on the skyline, overshadowed by taller more distant features, or to be ignored in the FOI summary as it does not form part of the FOI's silhouette. The FOI skyline summary is calculated from the ratio of points on the skyline over all visible points which are not assigned a null value.

```
Method CalculateTargetVisibility(ObserverPoint,TargetPoint)
TargetRatio = Ratio of elevation change over distance from Observer to Target

# Ray returns ordered set of points from observer to target (not inclusive)
FOR EACH TestPoint on RayofPointsFromObserverToTarget
   CurrentRatio = Ratio of elevation change over distance
            from Observer to TestPoint

 IF CurrentRatio > TargetRatio
      RETURN #target not visible - so return to main routine

 # Close horizon locations – inside and outside FOI boundary
 IF TestPoint Is INSIDE the FOI Polygon
   IF (CurrentRatio > MaxRatioInsideFOI)
      MaxRatioInsideFOI = CurrentRatio
      MaxPointInsideFOI = TestPoint
 ELSE
   IF (CurrentRatio > MaxRatioOutsideFOI)
      MaxRatioOutsideFOI = CurrentRatio
      MaxPointOutsideFOI = TestPoint

END FOR LOOP

# Calculate target interception elevations for close horizons
Intercept_Outside = Intercept Elevation at the Target for MaxRatioOusideFOI
Intercept_Inside = Intercept Elevation at the Target for MaxRatioInsideFOI

# Calculate vertical extent of target column is showing
IF MaxRatioOutsideFOI>MaxRatioInsideFOI
      TargetShowing = TargetElevation - Intercept_Outside
ELSE
      TargetShowing = TargetElevation – Intercept_Inside

# Calculate clearness index
IF Intercept_Outside>Intercept_Inside
   Clearness_Index = (TargetElevation – Intercept_Outside)
            / (TargetElevation - Intercept_Inside)
ELSE
   Clearness_Index = 1.0

# Facade area
Facade_Area = TargetShowing*TargetColumnWidth

# Perceived area
Perceived_Area=Facade_Area*ViewingAngleFactor*DistanceFactor

RETURN Facade_Area,Perceived_Area,Clearness_Index
```

```
Method CalculateTargetOnSkyline(ObsPt,TarPt)
OnSkyline = True #assumes the point is on the skyline until proven otherwise
TargetRatio = Ratio of elevation change over distance from Observer to Target

# Define search point on horizon
HorizonPoint= A point 50km behind target along bearing from Observer to Target

FOR EACH TestPoint on RayofPointsFromTargettoHorizon
   CurrentRatio =
         Ratio of Elevation change over distance from Observer to TestPoint

   # if the test point is visible then determine if in same FOI or not
   IF (CurrentRatio >= TargetRatio
          IF (TestPoint Is INSIDE FOI Boundary)
             OnSkyline=Null  # not part of silhouette outline – ignored in summary
             EXIT FOR LOOP
      ELSE
             OnSkyline=False  # overshadowed so not on skyline
         EXIT FOR LOOP

   CurrentRayElevation= ObserverElevation
                  +(TargetRatio*Distance from Observer to TestPoint)

   # Stop checking if current tests are above highest elevation in DSM
   IF CurrentRayElevation > Maximum Elevation in DSM
      EXIT FOR LOOP

END FOR LOOP

RETURN OnSkyline
```

3.6 Implementation and Evaluation

The five metrics, described in Section 3.5, were tested in a number of synthetic and real world scenarios. The trials were conducted by implementing a client-server architecture whereby the server carries out the visual exposure modelling in parallel across a number of CPUs, returning results to the client. The client is typically a GPS equipped mobile phone communicating across a wireless network (e.g. WiFi, 3G), but for testing purposes this may be substituted with a map based interface which simulates the mobile's location and orientation output.

3.6.1 Synthetic Surface Examples

In the first synthetic example a flat plain is considered, with a single FOI. The viewing angle and distance effects on perceived area are ignored, and only the total façade area visible is calculated from the surroundings. The FOI dimensions were set to be 150m long by 130m wide, and 100m tall. Although the structure is never blocked by nearby features the visible area changes as more than one side can be seen, as shown in Figure 3.11.

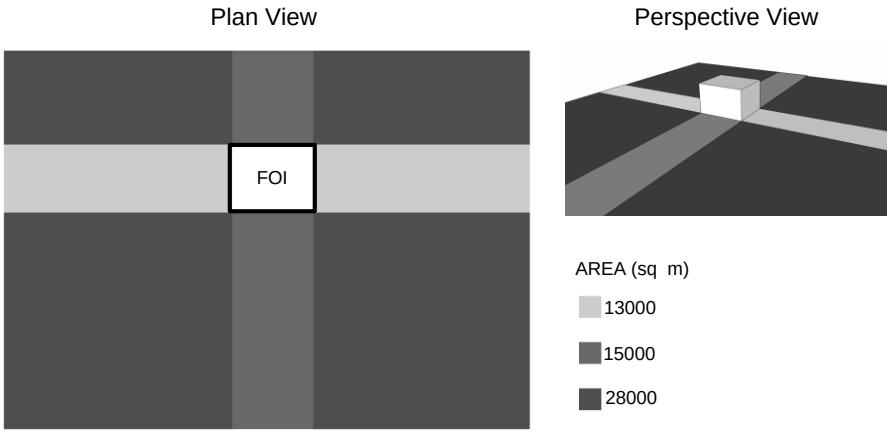


Figure 3.11: Visible Area of FOI on Plain - Excluding Distance and Viewing Angle

When distance and viewing angle factors are introduced the results are dramatically altered, as shown in Figure 3.12. The perceived area changes at the squared rate of the change in distance and therefore when an equal weighting is assigned to each factor the effects of distance are more pronounced.

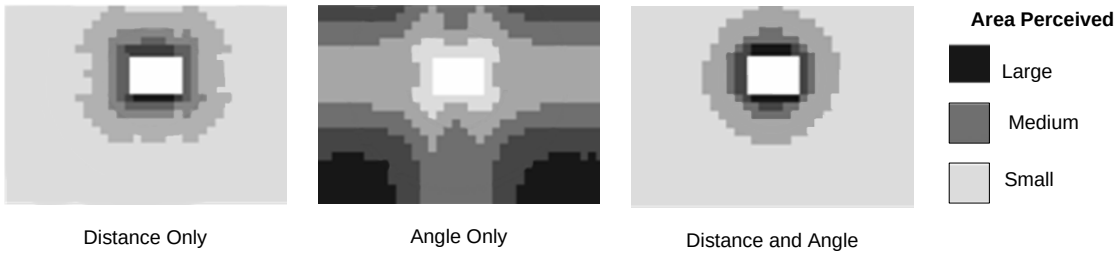


Figure 3.12: Distance and Angle Effect on Perceived Area

In this synthetic example the results show that no matter where the observer moves the entire height of the structure is always visible giving a clearness index of 100%. The object is also calculated to be 100% on the skyline, indicating no distant buildings overshadow it.

Next a low wall is introduced which partially obscures the view of the FOI from some locations. The effect of this is that the minimum visible elevation for the FOI changes as the user approaches. The area visible reflects this, although the horizontal field of view is unchanged. When close to the wall the FOI is not visible at all. Figure 3.13 show the results of these calculations, excluding distance and angle effects.

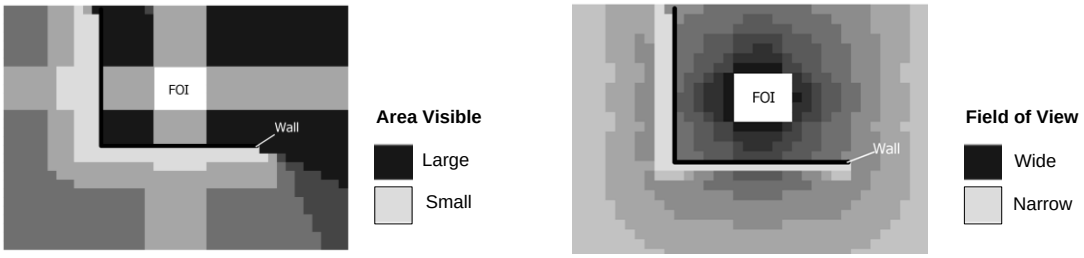


Figure 3.13: Synthetic Example with Wall - Calculated without Distance and Angle Effects

3.7 Real World Examples

The next trials were conducted in a real world scenario, based on a 1 metre resolution DSM and DTM created from a LiDAR dataset for Christchurch, New Zealand. A vector layer was used to define the boundary for each FOI, and all of the enclosed cells are considered as targets.

3.7.1 Visibility Calculation Trials in an Urban Region

A complete set of visibility metrics were calculated from a location in the city centre, from which point a number of key landmarks are visible. A panoramic image taken from this location is shown in Figure 3.14, and the calculated visibility results in Table 3-1.

Figure 3.14: Cathedral Square, Christchurch, NZ

Table 3-1: Visibility Results for Cathedral Square, NZ

FOI	Targets Visible (%)	FOV (degrees)	Façade Area (sq m)	Perceived Area (sq cm)	Clearness (%)	Skyline (%)	Lowest Visible Elevation (m above DTM)
		[Section 3.5.1]	[Section 3.5.2]	[Section 3.5.3]	[Section 3.5.4]	[Section 3.5.5]	
A	7	32	789	1148	47	97	13
B	28	37	516	1073	64	22	0
C	25	13	1209	396	49	100	28
D	20	20	426	378	17	11	4
E	10	24	254	78	54	46	2
F	47	19	1095	911	77	71	1
G	67	18	227	213	45	11	1
H	51	8	225	95	33	99	17
I	16	6	1076	197	58	87	4
J	17	42	860	8503	62	90	0

The *Targets Visible* metric gives an indication of how much of the building can be viewed from this location. Buildings A and E receive the lowest values, indicating that only a small portion of the entire FOI can be seen. Figure 3.15 shows a map of the targets visible for each FOI, confirming that the majority of Building A and Building E are hidden from view. This metric is useful for giving an indication of how much of the structure is out of sight.

The *FOV* column relates to the horizontal angle occupied by the FOI in the viewer's vision, a factor of its size and the viewing distance. In this example Buildings J and B have the highest values, while the distant skyscraper I has the smallest FOV. An apparently high FOV has been calculated for Building E, which is due to the section of the FOI visible north of Building F as seen in Figure 3.15. A better way to establish the size of a structure is to calculate the visible area, as illustrated next.

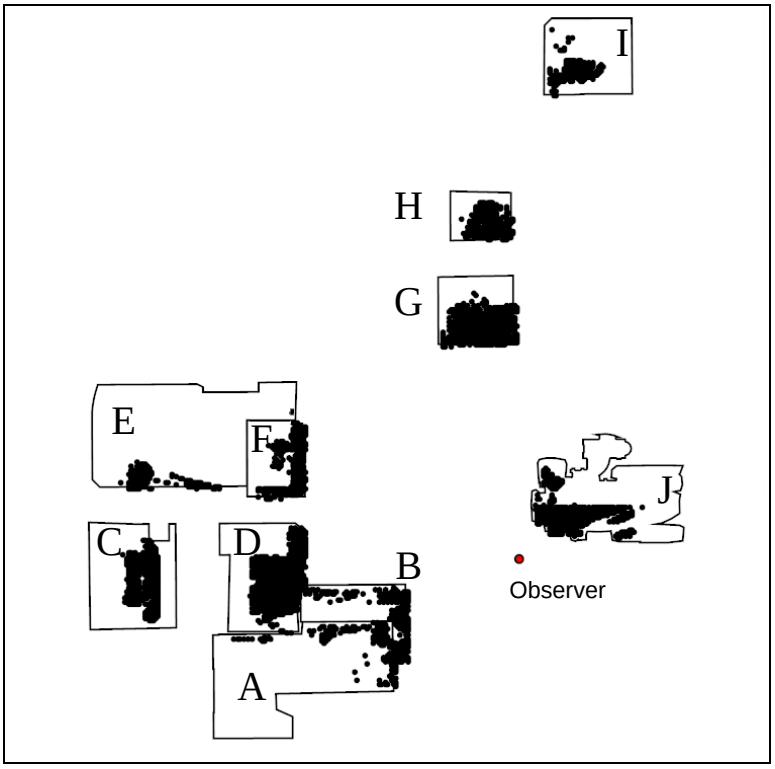


Figure 3.15: Map of Visible Targets for FOIs in City Centre

Table 3-1 has two columns for visible area. The first shows the *façade area* visible irrespective of distance or viewing angle, and the second takes these factors into account resulting in a *perceived area*, as described in Sections 3.5.2 and 3.5.3. These results show Buildings C, F, I to be the largest physical structures (from the *façade area* column) but from this viewpoint the largest perceived Buildings are J, A, B and F. The order of building dominance as ranked by perceived area is shown in Figure 3.16, and it is notable that Building J is much more prominent, in respect of its perceived visible area than the other FOIs. This is largely due to the proximity of J, however there are also situations where a more distant FOI is ranked above closer ones as a result of a larger area being visible, as in the case of FOI A which is ranked above B despite being further away. The same applies to buildings C and D, with C ranking higher than D despite its greater distance from the observer.

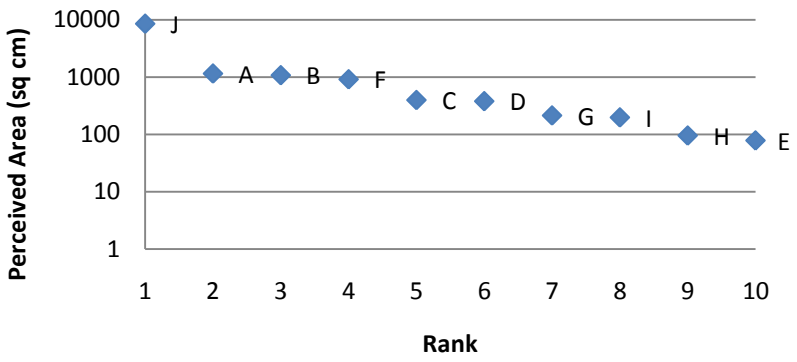


Figure 3.16: Features of Interest Ranked According to Perceived Area

The *lowest visible elevation above the DTM* column shows if the base of the building can be seen, or as in the case of buildings A,C and H are totally behind closer structures.

The clearness index represents how much of an FOI is obstructed by closer objects. In most cases the values reflect the real world experience, such as Building A being assigned a lower score than Building B which is in front of it. However Building F is given a rather high value, when in reality it is obscured by a number of trees. The reason was found to be that the LiDAR dataset has failed to capture the full width and complexity of the tree shapes, and therefore their effect is underreported in the results. This should be less of an issue with more recent LiDAR capturing equipment (Reitberger *et al.* 2006), which can capture detailed vegetation profiles. Apart from this anomaly buildings B, J, and I are assigned appropriate values which indicate they are clearly visible. A more detailed examination of the *clearness index* is given in Section 3.7.3.

Finally, the *skyline* metric uses the method outlined in Section 3.5.5 to quantify how much of the FOI makes the horizon, rather than being overshadowed by more distant taller buildings. The results show that buildings D,G and B are mostly overshadowed by taller more distant buildings, while C, H, and A are on the skyline. The parts of the FOI on the skyline may be mapped as shown in Figure 3.17.

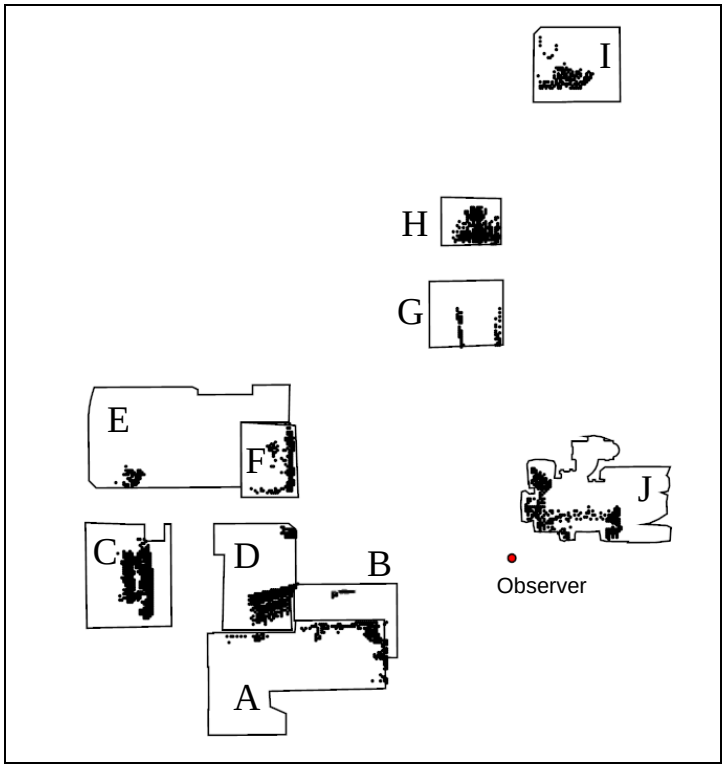


Figure 3.17: Map of the Skyline Targets for each FOI

3.7.2 Studying the Change in Metric Values for a FOI from a New Viewpoint – Case Study 1

The FOI visibility metrics are dynamic, modelling the user's viewpoint as they move through urban space. In this case the calculations for Landmark J are run from a new observation point, as shown in Figure 3.18. From this location the left side of the building is overshadowed, as reflected in the *skyline* value which has fallen from 90% (Table 3.1) to 48% (Table 3.2). Additionally the *targets visible* and *area* values have increased as we are now able to see a greater extent of the structure. The dramatic increase in *perceived area* is a result of the increase in area visible, viewing angle and distance. Figure 3.19 shows a map of the location of the visible targets, denoting those which are on the skyline, and also showing the position of the distant horizon behind the overshadowed targets. Finally notice that the *clearness index* is now higher as there are no trees in between the observer and FOI from this new viewing angle.

Figure 3.18: Visibility of Building J

Table 3-2: Visibility of Building J from an Alternative Viewing Angle

FOI	Targets Visible (%)	FOV (degrees)	Façade Area (sq m)	Perceived Area (sq cm)	Clearness (%)	Skyline (%)	Lowest Visible Elevation (m above DTM)
		[Section 3.5.1]	[Section 3.5.2]	[Section 3.5.3]	[Section 3.5.4]	[Section 3.5.5]	
J	21	79	2079	13832	96	48	1

Figure 3.19: A Map of Visible Targets, Skyline Points, and Distant Horizon Locations

A visible target cell will either make the skyline itself, or be overshadowed by a more distant structure.

The location of interception with the distant structure may be calculated and displayed.

3.7.3 Studying the Change in Metric Values for a FOI from a New Viewpoint – Case Study 2

A further study was carried out to compare visibility results on a single building from a number of viewing angles and distances. As can be seen from Figure 3.20 at viewpoint K only the top of the FOI is visible across rooftops. Table 3-3 shows the modelled results match this real world experience, indicating the *lowest visible elevation* for the building is 17m above the DTM. The percentage of *targets visible* does not change when the FOI is viewed from a closer viewpoint at the same viewing angle (viewpoint L). However the *lowest visible elevation* value falls to 4m, and the visible *area* increases from 927 sq.m, to 1,887sq. m, indicating the user has a much better view of the FOI (this *area* figure is irrespective of viewing distance). The *clearness* score also reflects this increasing from 22% to 73%. At viewing location M the entire side of the FOI is visible, behind a low vegetation layer. From this viewpoint the user has the greatest exposure to the building shape and highest number of visible targets. However by moving to the front of the building the observer has a clearer view unobstructed by vegetation, as a result the *clearness* value increases from 77% to 87%. The *visible target* count falls to 8% as the observer is only able to see a small proportion of the building, as also indicated by the area visible results now at 879 sq. m, rather than 1,478 sq. m that could be viewed at location M.

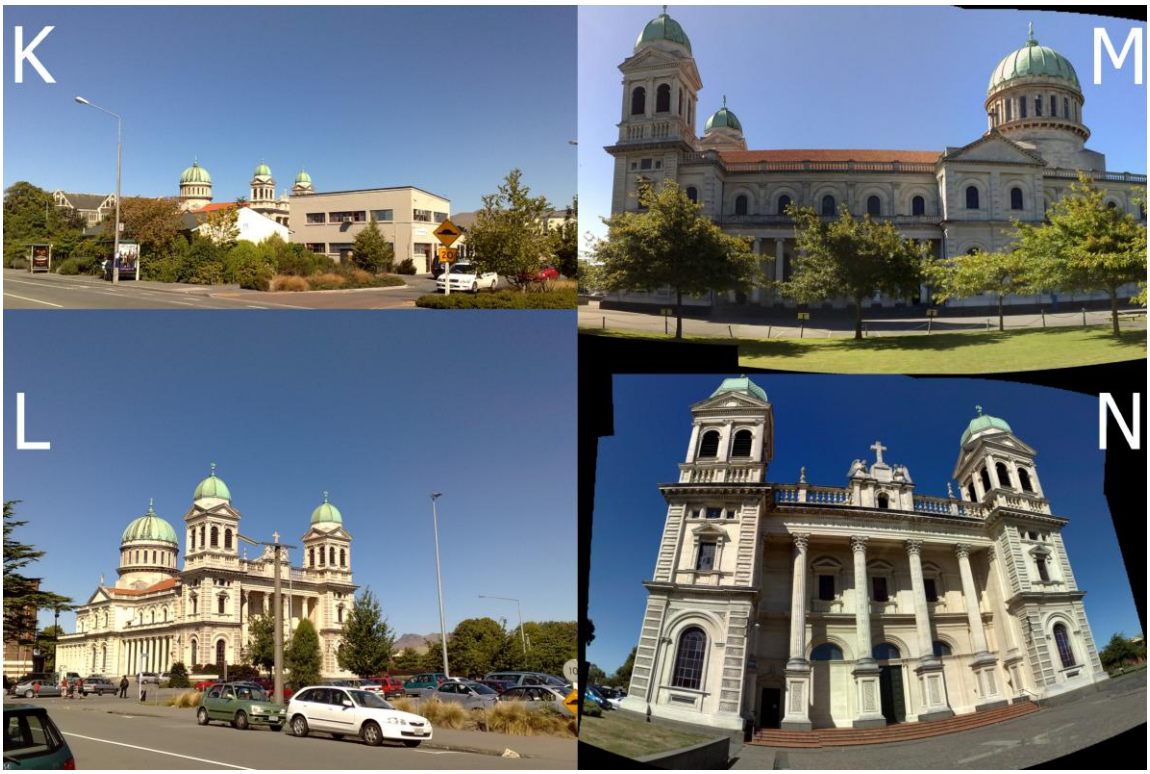


Figure 3.20: Visibility Metrics for a Single Building from Different Angles and Distances

Table 3-3: Visibility Results for Single FOI from Multiple Viewpoints

FOI	Targets Visible (%)	FOV (degrees) [Section 3.5.1]	Façade Area (sq m) [Section 3.5.2]	Perceived Area (sq cm) [Section 3.5.3]	Clearness (%) [Section 3.5.4]	Skyline (%) [Section 3.5.5]	Lowest Visible Elevation (m above DTM)
K	31	13	927	145	22	100	17
L	31	25	1887	2526	73	98	4
M	40	90	1478	17624	77	98	3
N	8	92	879	6191	87	94	2

3.7.4 Mapping Visual Metrics

In the final real world test the metrics are mapped back into the surrounding space, revealing visual exposure patterns for each visual property, as shown in Figure 3.21. In this example the visible number of targets increases with distance from the FOI, as does the Façade Area visible. These are measures of how much of a building structure is visible irrespective of distance, and as the observer moves away from the structure so more of it comes into view. However the Field of View and Perceived area maps show a decrease as the observation distance increases.

Maps may also be derived from the results, for example using flow direction functions to calculate in which direction the observer should move for a clearer view of the FOI. The direction to move for a more limited view may also be calculated, and it should be noted this is not an inverse of the clearer view map as explained in Section 3.5.4.

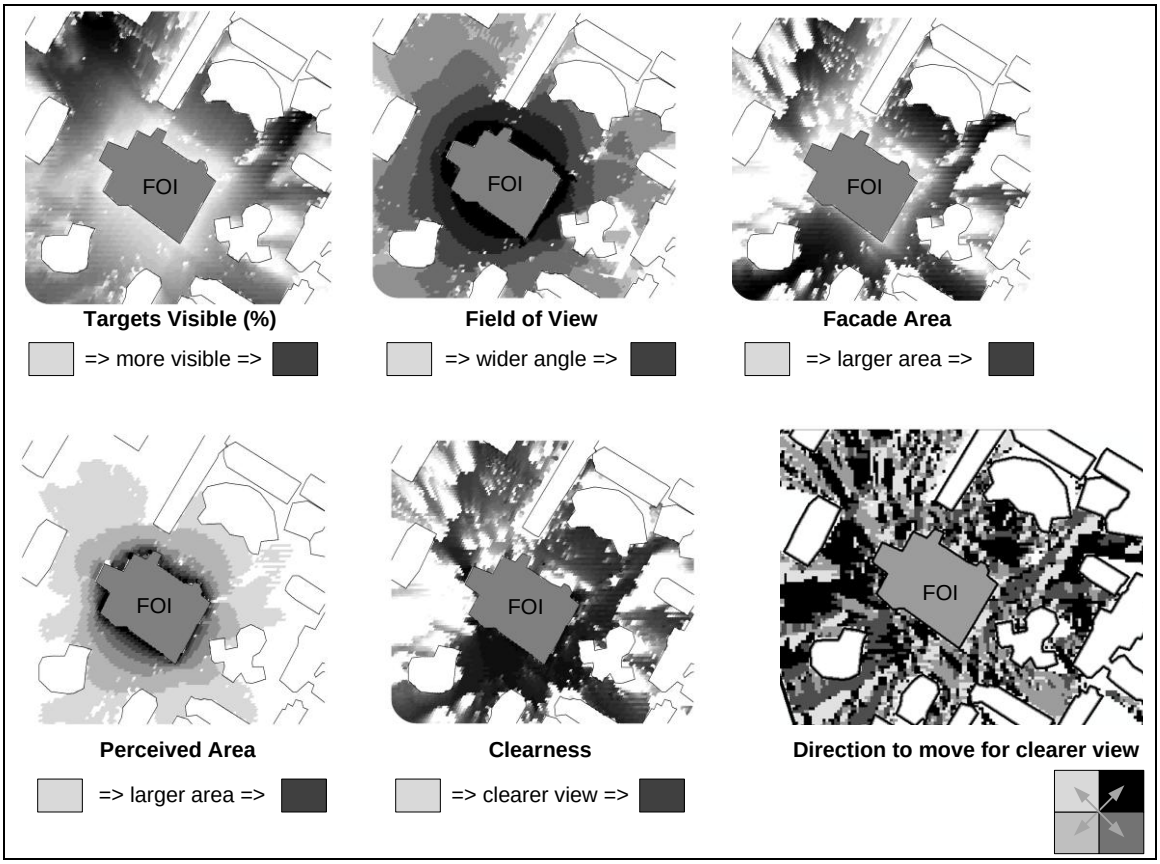


Figure 3.21: Map of Visual Metrics

The dominance of an FOI within a region may be determined by ranking it according to one of the visibility metrics. This is achieved by totalling the calculations from many observation locations and linking back the results to each FOI. An example using the skyline metric within a university campus study region is shown in Figure 3.22. Here the total skyline percentage was calculated from every square metre in the study area, for each FOI. The skyline percentages from each observation point were then summed for each FOI, giving a single skyline value for the FOI for the entire region. In this example the FOIs with the highest rankings (i.e. most often on the skyline and not overshadowed) are the tallest structures (Figure 3.22 – FOI 3), or those with the largest floor area (Figure 3.22 – FOI 1), often located near the outskirts of the campus. The lowest rankings occur where a low-rise building is surrounded by much taller structures. This technique may be incorporated into the process of automatically identifying the most significant landmarks within an urban region.

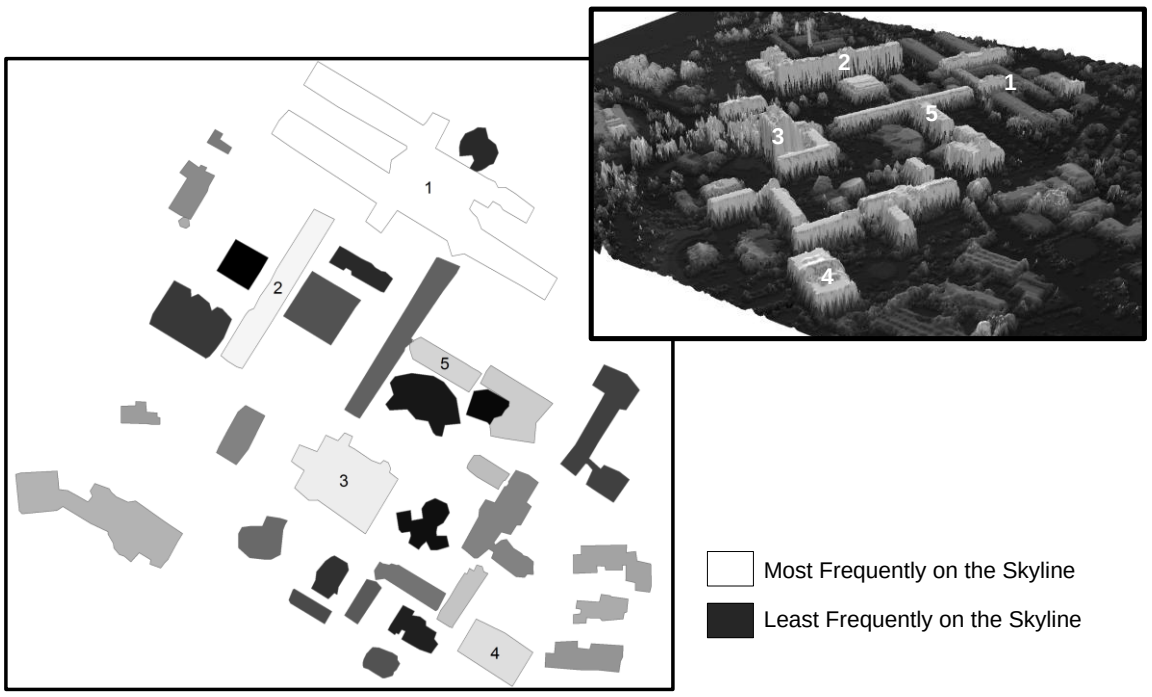


Figure 3.22: FOIs ordered by how frequently they are on the Skyline when viewed from the surroundings
The top 5 most frequent skyline FOIs are labelled, and also displayed in a 3D view generated from a LiDAR dataset.

3.8 Conclusion and Future Work

Current visibility modelling tools in GIS provide the functionality to estimate how much of the surrounding region is visible from an observation point. This is particularly useful for measuring the impact of planned developments on the surrounding region, or the feeling of openness in built-up environments (Fisher-Gewirtzman and Wagner 2003, Yang *et al.* 2007). Typically isovist models (Benedikt 1979) are used to describe the extent visible in urban environments, while viewsheds (Fisher 1991, Fisher 1995, Wang *et al.* 1996, De Floriani and Magillo 2003, Cooper 2005) have been applied in rural regions to calculate which portions of the countryside may be seen by an observer.

However only limited research exists for modelling the 'visual exposure' (Llobera 2003) of a feature. This is an estimate of how much of a feature may be seen from the surrounding terrain by modelling its visual properties, such as the vertical field of view occupied, from every neighbouring location, such that an observer may be instructed in which direction to move to attain a clearer view of a landmark.

The research presented here uses 2.5D GIS datasets to model FOI profiles, such that façade areas may be calculated. This is achieved through an extended LOS model which considers the most prominent features in front of the FOI. Each FOI cell is considered as a vertical column between DTM and DSM models, and values for visible area are aggregated for all the visible

cells within a designated FOI boundary to generate object level summaries. These summaries are intended to quantify the visibility of FOIs, allowing future LBSs to add aspects of feature visibility into the array of searching and filtering tools available, and therefore to build a greater contextual understanding of the user's current situation. The visibility algorithms take into account the elevation profile of the FOI and the surrounding city, and are able to deduce if an FOI is out of sight, clearly visible, partially obscured, or on the skyline. FOIs may be ordered according to their perceived visible area by taking into account the building's visible extent, viewing distance and viewing angle which may be in turn used by LBSs to prioritise information delivery.

A number of trials were carried out on a synthetic DSM to establish that the algorithm's function, prior to running real world demonstrations. Adaptations were required so that concepts such as 'on the skyline' and 'clearness' indexes could be calculated successfully. The final model was able to produce results which reflect the experiences of the observer, quantitatively expressing the degree to which FOIs are visible, or obscured from view.

The algorithm is computationally intensive, which restricts its use on mobile devices directly. Two approaches which may be used to overcome this limitation are to pre-render the results for a set of FOIs in a study region (e.g. city, campus), or to provide a client-server architecture.

The pre-rendered approach works by first processing the visibility results from all possible locations within a designated study area to build a result cache. This requires a considerable computation facility but is largely aided by considering only those locations which are accessible to pedestrians outside of buildings, thereby reducing the number of observation points. In addition the algorithm may be run in parallel across a number of CPUs, as each target visibility test is independent. The FOI summaries do however require the results to be brought together on a single processor. Once the cached dataset is available it may be readily used in real time on current mobile platforms, which need only look up the appropriate results for a location. The main disadvantage of this approach is that any changes to the DSM or FOI database require reprocessing and cache distribution.

The client-server approach requires no advanced processing and provides real time server side support for a mobile client. Here the mobile application returns the location and orientation information to the server, which runs the visibility model in parallel across a number of CPUs to return the FOI visibility information in near real time. The main advantage of this method is that the DSM and FOI database may receive regular updates without the need to distribute this to clients. The main disadvantage is that sending data over wireless networks imposes an ongoing usage cost, which hopefully will reduce with the advent of WiMax (Ohrtman 2005, Patton *et al.* 2005) and other regional wireless network data solutions.

The processing time is largely dependent on the CPUs available, the number of FOIs in the database, and the resolution of the DSM. To improve the processing efficiency an additional step

may be added whereby the FOI database is filtered very quickly to determine which are visible, and therefore should be modelled in full. This is done by calculating the skyline polygon around the observer, and using this to find the intersecting FOI polygons. The skyline polygon requires that only a few thousand points are sampled, resulting in a significant performance increase in the overall algorithm, as the number of FOI targets which can be removed from the scan list will be significantly more than this.

There are a number of improvements which can be made to the algorithm including consideration of partial visibility through vegetation and methods to handle position uncertainty. These would both introduce a probability to the result, based on the shape and density of vegetation and accuracy assessment of the observer location. The latter should become less of an issue with the advent of improved tracking solutions to assist GPS in difficult environments, such as Inertial Measurement Units (Radoczky 2007).

Additional factors which influence the identification of FOIs, such as their use, shape, texture, colour, materials, and architectural design would be useful additional metrics to include in calculating a landmark dominance index, and to assist in FOI identification in urban scenes. Establishing the relationship between these metrics and the user's perceptions of landmarks will require user trials and should form the basis of an algorithm which can automatically establish, from any location, the most significant buildings, and those most easily identified.

3.9 Acknowledgments

The authors would like to thank Dr William Mackaness for comments on a draft of the chapter, and the Christchurch City Council for use of their LiDAR dataset. The research was supported by funding from the University of Canterbury, New Zealand.

Chapter 4: Incorporating Vegetation into Visual Exposure Modelling in Urban Environments

4.1 Summary

Visual exposure modelling establishes the extent to which a nominated feature may be seen from a specified location. The advent of high resolution Light Detection and Ranging (LiDAR) sourced elevation models has enabled visual exposure modelling to be applied in urban regions, for example to calculate the field of view occupied by a landmark building when observed from a nearby street. Currently visual exposure models access a single surface elevation model to establish the lines of sight between the observer and landmark feature. This is a cause for concern in vegetated areas where trees are represented as solid protrusions in the surface model totally blocking the lines of sight. Additionally the observer's elevation, as read from the surface model, would be incorrectly set to the tree top height in those regions. The research presented here overcomes these issues by introducing a new visual exposure model which accesses a bare earth terrain model, to establish the observer's true elevation even when passing through vegetated regions, a surface model for the city profile, and an additional vegetation map. Where there is a difference between terrain and surface elevations the vegetation map is consulted. In vegetated areas the line of sight is permitted to continue its journey, either passing under the canopy with clear views, or partially through it depending on foliage density, otherwise the line of sight is terminated. This approach enables landmark visual exposure to be modelled more realistically, with consideration given to urban trees. The model's improvements are demonstrated through a number of real world trials, and compared to current visual exposure methods.

4.2 Introduction

Visibility models may be used in the planning phase of developments, for example to calculate the visual impact of a new wind farm, or to find the most hidden path for a motorway extension. They have also been used in Location Based Services (LBS) to calculate what a user may see, and therefore enable context relevant data filtering, and customised content delivery (Bartie and Mackaness 2006). In all applications model fidelity has improved as more comprehensive digital datasets have become available, yet little consideration has been given to how vegetation may be accommodated within these models. Vegetation is a special case because its impact may be seasonal, and the observer is able to see partially through the canopy layer, as well as clearly underneath it. This research focuses on how a vegetation map may be incorporated into a

visibility model able to report the visual exposure of nominated landmarks in urban environments.

There are two main types of visibility models, isovist and viewshed models. Isovist modelling is suited to urban areas, establishing the expanse of continuous visibility around an observer by accessing building footprint polygons (Tandy 1967, Benedikt 1979), while viewshed analysis (Tandy 1967, Lynch 1976) using terrain models is more commonplace in rural regions. The introduction of Light Detection and Ranging (LiDAR) techniques enables high resolution georeferenced elevation datasets to be captured across large areas in minimal time suitable for producing Digital Surface Models (DSMs) in urban environments (Palmer and Shan 2002, Rottensteiner and Briese 2002). As a result viewshed calculations are now feasible within urban regions, as both topography and built form are captured within the surface model.

An issue which remains is how to handle non-surface urban vegetation, which appear in the DSM as solid blockades protruding from the ground. In this research a basic vegetation map is incorporated to improve the model's performance when calculating the visibility of landmark buildings in an urban environment. The model accesses a Digital Terrain Model (DTM) to establish the observer's elevation, and a vegetation map to distinguish surface vegetation from buildings in the DSM. From this an adapted Line of Sight (LoS) algorithm is able to establish under canopy and through canopy views, summarising the visibility for each landmark as a number of metrics (e.g. façade area visible). The chapter begins with a general introduction to visibility modelling and an outline of the existing approaches for including vegetation within the visibility model (Section 4.3). This is followed by an explanation of the solution implemented here (Section 4.4), and a number of real world trials to demonstrate the benefits of the enhanced model (Section 4.5).

4.3 Visibility Modelling Background

Geographic Information Systems (GISs) are able to carry out visibility calculations by accessing a DSM which stores elevation values for a region. The visibility calculations determine which locations can be connected by a straight line without being interrupted by the terrain (Franklin and Ray 1994). A good overview of the algorithms, storage formats and techniques to calculate visibility may be found in De Floriani and Magillo's article (2003). Much of the research has focused on improving algorithm performance (De Floriani *et al.* 2000, Rana and Morley 2002, Rana 2003, Ying *et al.* 2006), which has become increasingly necessary since the introduction of high resolution DSMs such as those sourced from LiDAR.

Modelling may be carried out on a point-to-point basis, or point-to-area as in the case of viewsheds (Fisher 1991, Fisher 1994a, Lee 1994, Fisher 1995). Viewshed results are normally stored as Boolean rasters which represent visible and hidden cells. They have been used to assist

in the siting of transmission masts, for calculating the visual impact of planned new developments (Fisher 1996), and for finding the most hidden routes from the surroundings (e.g. pipelines, military) or best views along a route (e.g. tourists) (Stucky 1998), among other things.

Caldwell *et al.* (2003) demonstrated how Complete Intervisibility Databases were able to answer new questions, such as giving an estimate of the percentage of a target visible, by summing the total number of visible cells within a defined zone. The introduction of the 'visual exposure' concept (Llobera 2003) established another new ability for visibility modelling, whereby a target's exposure may be mapped across a terrain. Here the model looks inwards at a designated feature, establishing how much of it may be seen from any surrounding location for a given visual property, such as the occupied horizontal field of view. Flow direction maps may be generated from these visual exposure maps to indicate in which direction an observer should move to attain a clearer, or more limited, view of a target.

The visual exposure model may be applied to urban features (e.g. a landmark building) by considering groups of cells, within a designated boundary, as a single entity. From this summaries may be produced which describe a number of visual properties (e.g. visible façade area) for the feature (Bartie *et al.* 2008). This differs from viewshed and isovist modelling as the specific focus is on calculating how much of a surface feature can be seen, rather than mapping the visibility of the terrain itself. However the results are subject to errors which result from the way vegetation is represented in 2.5D DSMs, as solid jagged topped barriers which block the LoS. Furthermore, when the observer is located amongst trees, the modelled view is calculated as if the observer were at tree top height, as read from the DSM. To improve this model a visibility tool has been developed that begins the LoS calculation using an observer height offset from the ground level as stored in a DTM, while considering the profile from the DSM, and where designated permitting the LoS to pass partially through the DSM according to its visual permeability, thereby accommodating under canopy and through canopy conditions. The next section describes the implementation in an urban environment for the purpose of modelling landmark building visibility, beginning with the background to current methods for including vegetation in visibility modelling.

4.3.1 Visibility Modelling Incorporating Vegetation Background

There have been a number of previous studies which incorporate vegetation within the visibility model. Dean (1997) introduced the concept of visual permeability, proposing that a ray of light could pass through a modelled canopy layer, with a linear attenuation proportionate to the distance travelled in the canopy space. Dean's model was updated by Llobera (2007a) who considered visual permeability as photons travelling through a medium following the principles outlined in Beer-Lambert's Attenuation law. The model considers a beam of light as it passes through vegetation, experiencing an exponential drop in photon numbers for each unit of

distance travelled, indicating the probability of viewing a region as determined by the spatial density and position of tree models on a landscape. The research demonstrates the possibility of the improvements to visibility modelling in rural regions, and highlights a number of the difficulties in sourcing high quality vegetation maps, and information on tree species for such a purpose. Llobera concludes that the research has not been tested empirically and to do so would require an area dominated by one tree type. There is no consideration for how the technique may be incorporated into visual exposure modelling for use in urban regions, nor how LiDAR sourced DSMs and DTMs may be combined with remotely sensed imagery to produce vegetation models, which is the work presented in this chapter.

A compendium of LoS algorithms (US Army Corps of Engineers 2004) used within commercial and military GIS software identified a number of applications which are able to model the view under vegetation canopy. These make use of a second surface model which stores the canopy base heights (Baer *et al.* 2005). While enabling under canopy views, concepts of partial visibility through the canopy layer and data collection are not discussed.

Other military models use Probabilistic Line-of-Sight (PLoS) techniques (Stanford *et al.* 2003) to calculate the ability of different sensors to view targets. In these models trees are represented as simple 'ice cream cones' with a number of parameters to define trunk widths, tree proportions, and foliage density. A similar approach used by Liu et al (2008) represents trees as intersected diamonds. From this the visible regions on hillsides, and those blocked behind vegetation, can be calculated. While these approaches are useful adaptations for rural scenarios they do not cater for feature visibility modelling in urban areas, where the output required not only includes which raster cells can be viewed but also an indication of the vertical extent of each cell visible so that façade area calculations may be performed.

This chapter presents research on raster based tree models generated by supplementing LiDAR DSMs with Quickbird remotely sensed imagery and a minimal ground survey. The intended use for this model is in urban regions, to establish which Features of Interest (FOI) are visible from given locations, for use in LBSs as a filtering method for determining the relevance of surrounding items.

4.3.2 Urban Visual Exposure Metrics Background

Before developing concepts further it is worth summarising the current urban feature visual exposure model. Visual exposure modelling quantifies how much of a facet may be seen from a given location (Llobera 2003). By using a LiDAR sourced DSM it is possible to calculate the vertical intercepts above the DTM for LoSs cast from an observer to FOIs (Figure 4.1a) enabling surface area calculations to be carried out for any FOI in a region (Figure 4.1b). The resultant "façade area" quantifies the surface area of a FOI which may be viewed from that location irrespective of viewing distance or angle, essentially establishing the size of the visible object.

An additional "perceived area" metric includes the distance and viewing angle factors. A more detailed explanation of the model and definitions of the metrics may be found in Bartie *et al* (2010).

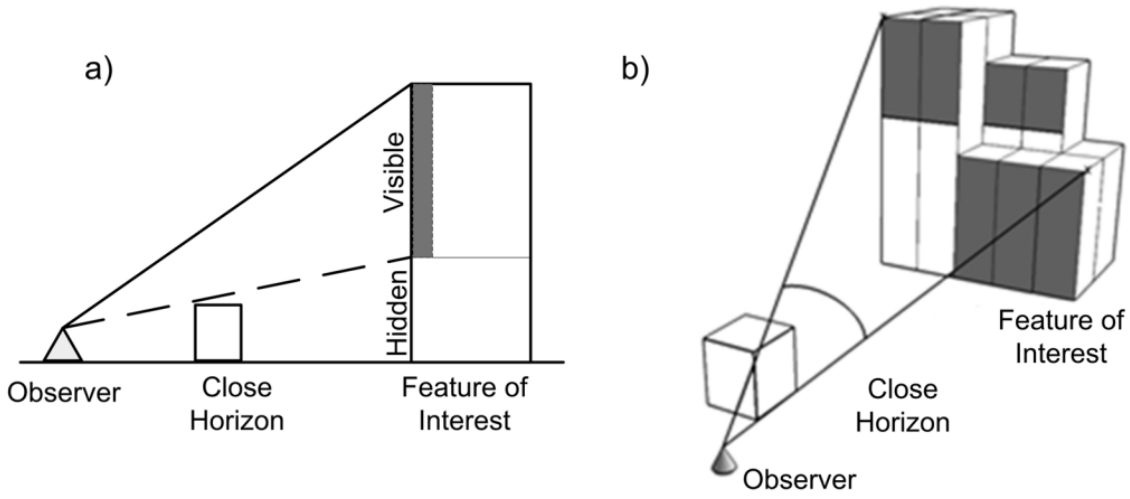


Figure 4.1: Visual Exposure Modelling: (a) Side View of modified LoS (b) Façade Area Visible

4.4 Accommodating Vegetation within a New Visual Exposure Model

One of the issues with modelling features in urban environments using LiDAR sourced DSMs is the representation of vegetation as solid walls due to the 2.5D data structure limitations. For any visibility studies this means that trees form total visual blockades completely blocking the line of sight. The issue is twofold, firstly trees may allow some partial visibility of distant objects through their canopy layer, and secondly the view either side of the trunk underneath the canopy layer is clear.

Furthermore if only a DSM is used then as the user moves into vegetated zones the view will be modelled as if the observer is standing on top of the vegetation (Figure 4.2a). One solution is to introduce a DTM bare earth model ensuring all observations begin at ground level, however if the observer is positioned within a vegetated area, then the LoS ray will be immediately terminated as it reaches the surrounding DSM cells which are higher than the observer's DTM value (Figure 4.2b). Therefore the visibility model requires further information to distinguish between the visual properties of the feature which occupies the gap between DTM and DSM. This can be accomplished by identifying the vegetated regions by means of a Normalised Differential Vegetation Index (NDVI), calculated using Near-Infrared and Red bands from high resolution Quickbird satellite imagery.

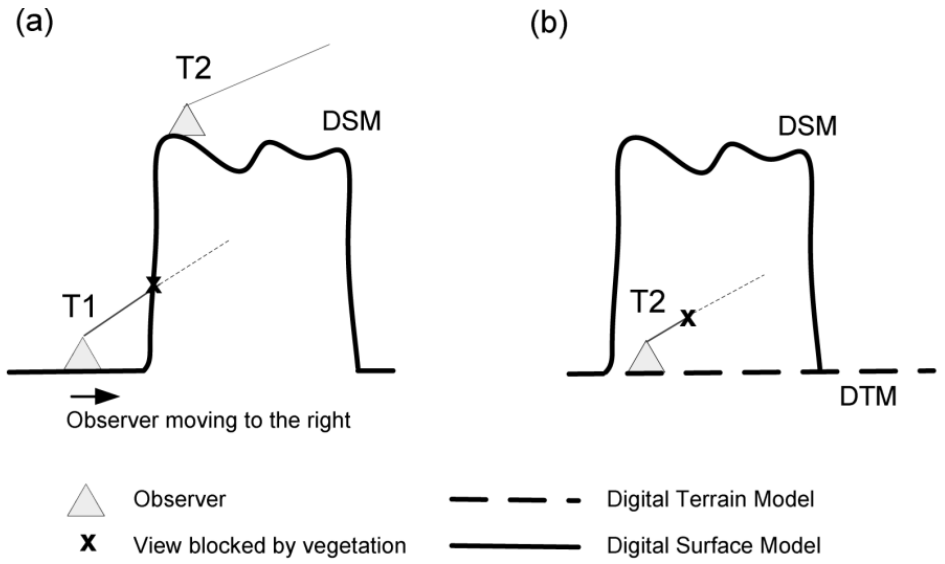


Figure 4.2: (a) A DSM created from LiDAR data showing an observer standing under a tree at T2. A viewshed generated from this location would consider the observer to be at tree top height. (b) Introducing a DTM to establish true ground height. See text for full commentary.

4.4.1 Building a Vegetation Map for use in Visibility Studies

The vegetation map is only required where a difference in elevation is noted between DTM and DSM, so that the visibility model may determine vegetated partially visible regions from solid structures (e.g. buildings, statues). These regions may be mapped using Quickbird imagery, which is supplied as 4 bands at 2.4m and a 60cm panchromatic band for image sharpening, to calculate the NDVI. Inevitably there will be radial distortion errors introduced when using Quickbird imagery, which can be overcome to some extent by buffering the vegetated areas by a few metres and then clipping them against the elevated zones where there is a difference between DSM and DTM values. To ensure that vegetation does not encroach into the FOIs a final clipping operation was carried out to remove any overlapping areas from within the FOIs polygons. An example of the process and resulting tree polygons are shown in Figure 4.3.

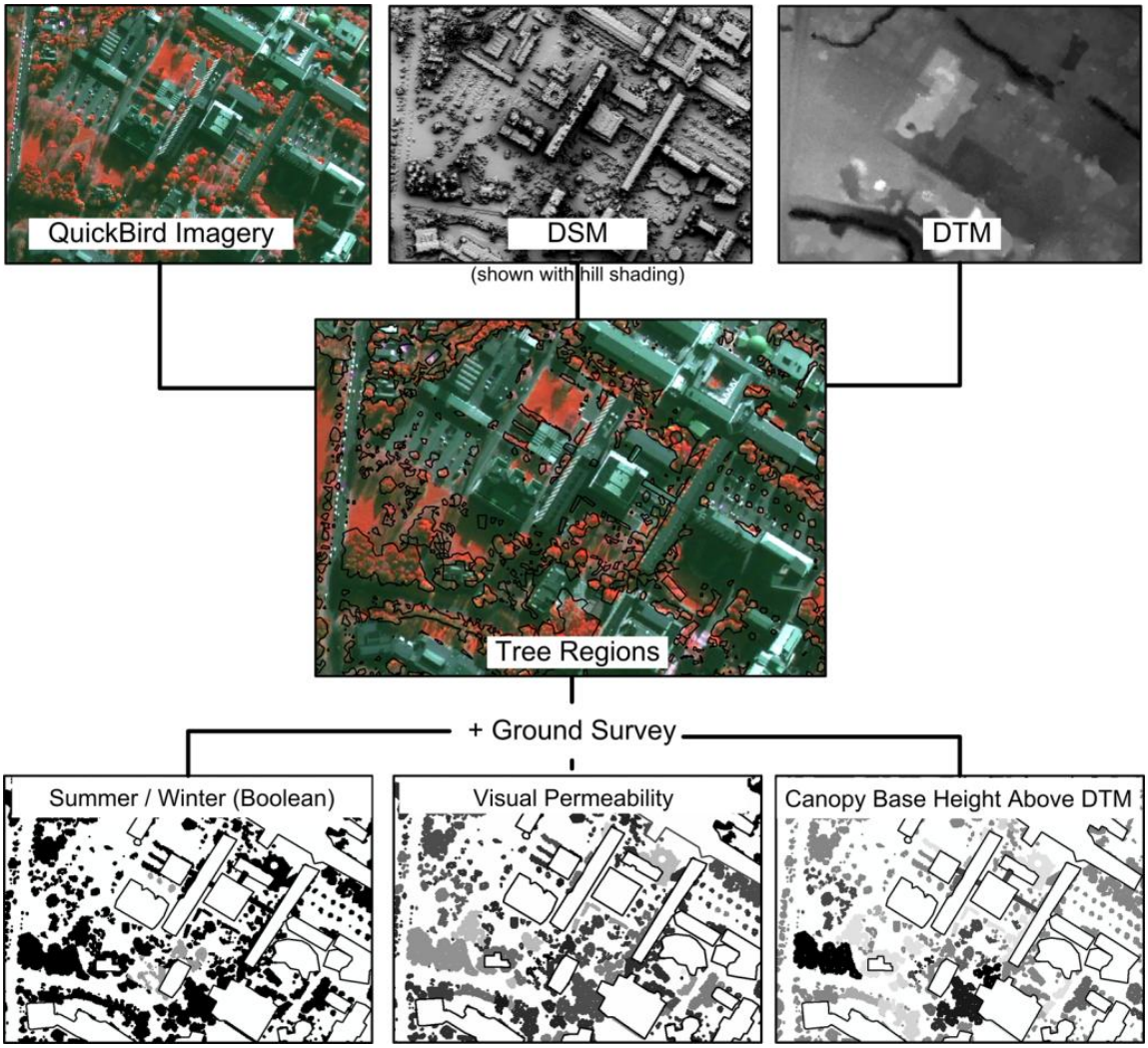


Figure 4.3: Tree polygons created from Quickbird Imagery NDVI and Difference in DSM and DTM elevations

The NDVI information is used to determine vegetated regions from non-vegetated to select between the DSM or DTM values along the LoS path. However more information on the tree leaf density and canopy base height is required to improve the visibility results. In urban areas traditional remote sensing tree classification techniques do not work well, as trees are found in low density highly heterogeneous groups, often over grassy regions (Xiao *et al.* 2004). Comprehensive vegetation density information may be collected using terrestrial side looking LiDAR, as shown in Figure 4.4. Where this is not available a ground survey will suffice in providing approximate density information for the vegetated zones. The supplementary information is stored as attributes against each vegetation cell defined by the NDVI process, ensuring that the additional information is correctly spatially registered to the DSM and DTM cells. Through studying a winter aerial image the deciduous trees were noted, and the position and approximate widths of tree trunks were added to the dataset.

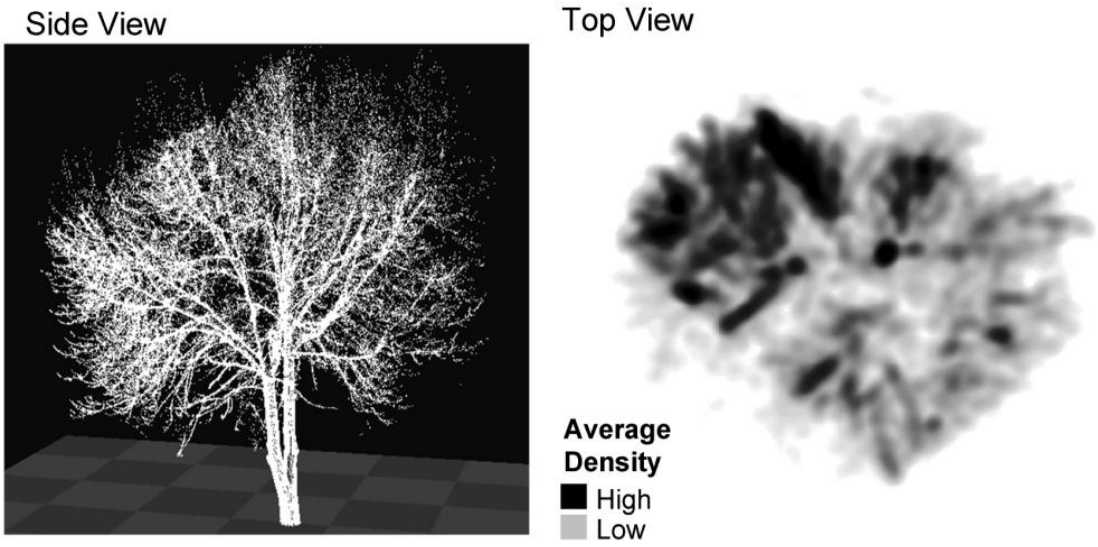


Figure 4.4: Terrestrial Side Looking LiDAR Tree Scan with Density Map

4.4.2 Implementation

The vegetation map supplies information on the position, visual permeability, and canopy base height so that the probability of the LoS passing through that region may be calculated. Canopy base height is recorded as the height above the DTM, and captured during the ground survey using a range finder. Generalised values for a tree or group of trees may be sufficient depending on the accuracy required in the modelled output, although variations at cell resolution as collected by terrestrial LiDAR scans may also be stored. Each tree cell is assigned a Boolean value indicating if it is affected by seasonal variation, so that the model may switch the visual properties when run in summer or winter months. The trunk locations are stored as cells with a visual permeability rate of 0.00, a canopy base height of 0 metres, and assigned no seasonal change values irrespective of tree species.

Ideally terrestrial side looking LiDAR surveys would be used to collect comprehensive vegetation density maps for urban areas, however this is not always feasible and therefore visual ground surveys are required. For this research tree permeability rates were collected for the designated vegetation regions, as defined by the NDVI map, using a visual assessment based on an image threshold technique, as shown in Figure 4.5. The permeability value gives an indication of the amount of light which can pass through the tree canopy from a side view, considering the tree to act as a filter. A value may be collected for the overall canopy, or separate readings may be taken for sections where the density is significantly different. This is then rasterised based on the tree depth from that sample point to give each cell in the vegetation map an approximate permeability rate, which describes the survival rate of a ray passing through that metre of canopy. For example if the collected permeability rate for a section of canopy was 53%, where it

was 3 metres deep, then the average cell permeability may be calculated using Llobera's (2007a) formula,

$$Ray\ survival = (1 - f)^m\ , \qquad (eq.\ 4.1)$$

where *f* is the ratio of tree matter to defined canopy space, and *m* is the depth the ray has passed through the tree. In this case the ray survival (i.e. 0.53) and tree depth (i.e. *m*=3) are known, therefore the average permeability rate (i.e. *1-f*) for a single cell may be calculated as 0.81 (i.e. $0.53^{\wedge 1/3}$). This approach gives an approximate value for the cell permeability as a function of canopy density, reflecting the variety of tree instances in urban scenes.

The LoS should be considered as a narrow column of light projected from the observer to the FOI, subjected to obstacles along the way. The permeability value for a vegetated region describes the division between solid tree and air for that cell, representing a vertical column between the tree top and canopy base. The higher the permeability rate the more of the LoS light column will reach the FOI, the lower the rate the more of the LoS will be blocked.

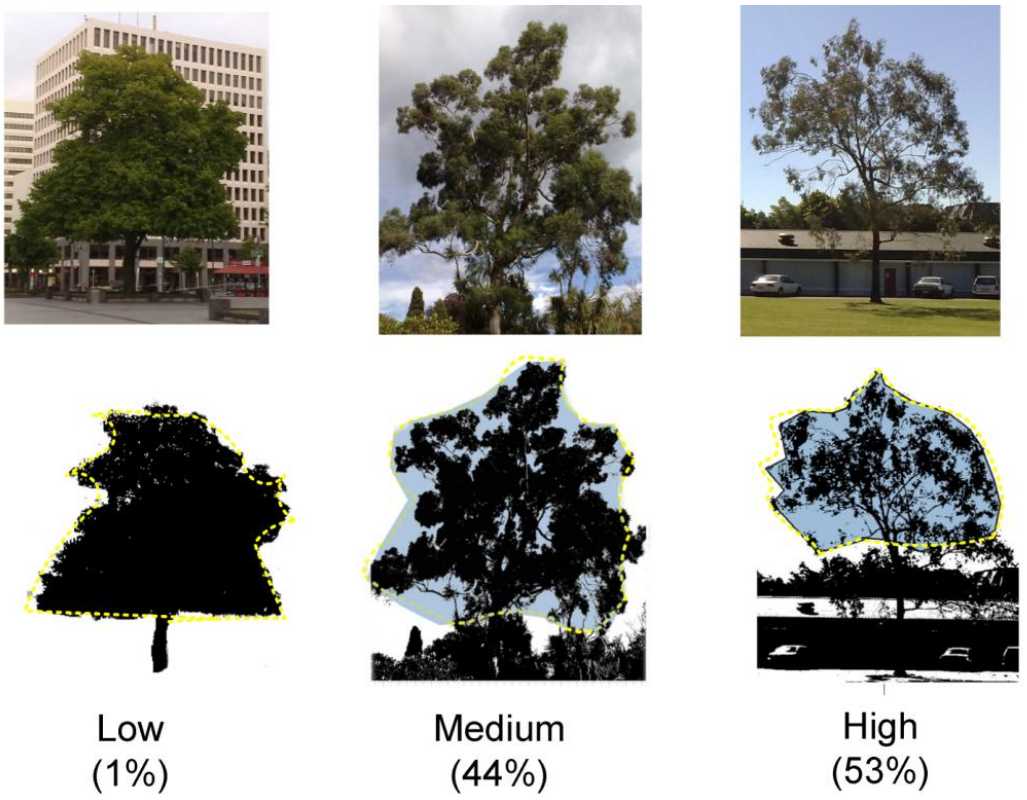


Figure 4.5: Examples of Tree Permeability (shown as a percentage)

To differentiate between the vegetation permeability values and those of the line of sight ray, we introduce the term *coverage index* to express the ray's current status along the path. The ray is assigned a starting *coverage index* of 100% and as it progresses towards the target is subjected

to varying rates of decay appropriate to the medium it is passing through. When passing through air an atmospheric rate is used, but when passing through vegetated zones the rate is read from the vegetation raster map (i.e. the vegetated cell permeability values). The ray is modelled as a 3D vector, along which a sample is taken at 1 metre intervals from the raster layers, as shown in Figure 4.6. At each sample point the height of the ray above DTM is compared to the vegetation canopy heights, to establish if the ray is passing under, through, or above the canopy layer. The model is therefore able to replicate the decrease in visibility associated with looking through deeper canopy sections, and as a result of the observer looking up through more layers of canopy upon approaching an FOI.

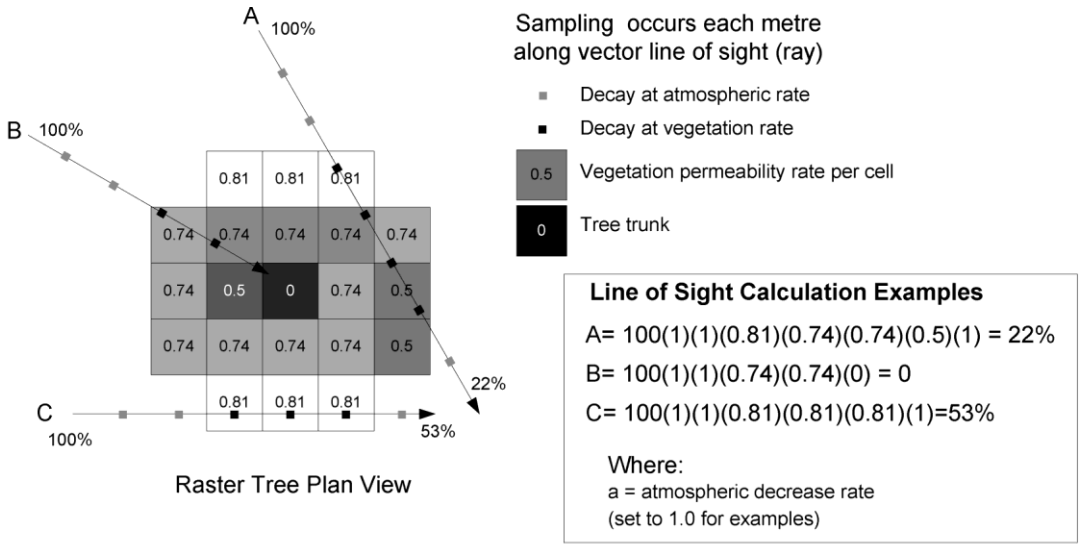


Figure 4.6: Example Rays Passing through a Raster Tree Model

Seasonal changes are not uniform across all trees, depending on their age, condition, and many other factors. While the model may accommodate this level of detail current data capture costs are restrictive, but it is hoped this will improve in the future as more sophisticated high resolution hyperspectral remote sensing techniques (Xiao *et al.* 2004), and ground based LiDAR surveys (Omasa *et al.* 2008) become more commonly available. For the purposes of this chapter the basic survey information and simple switch between summer and winter mode is sufficient to demonstrate the modifications to the algorithm. Figure 4.7 shows how the supplementary information is used within the vegetation model, storing information about tree canopy base height and permeability as additional pixel attributes.

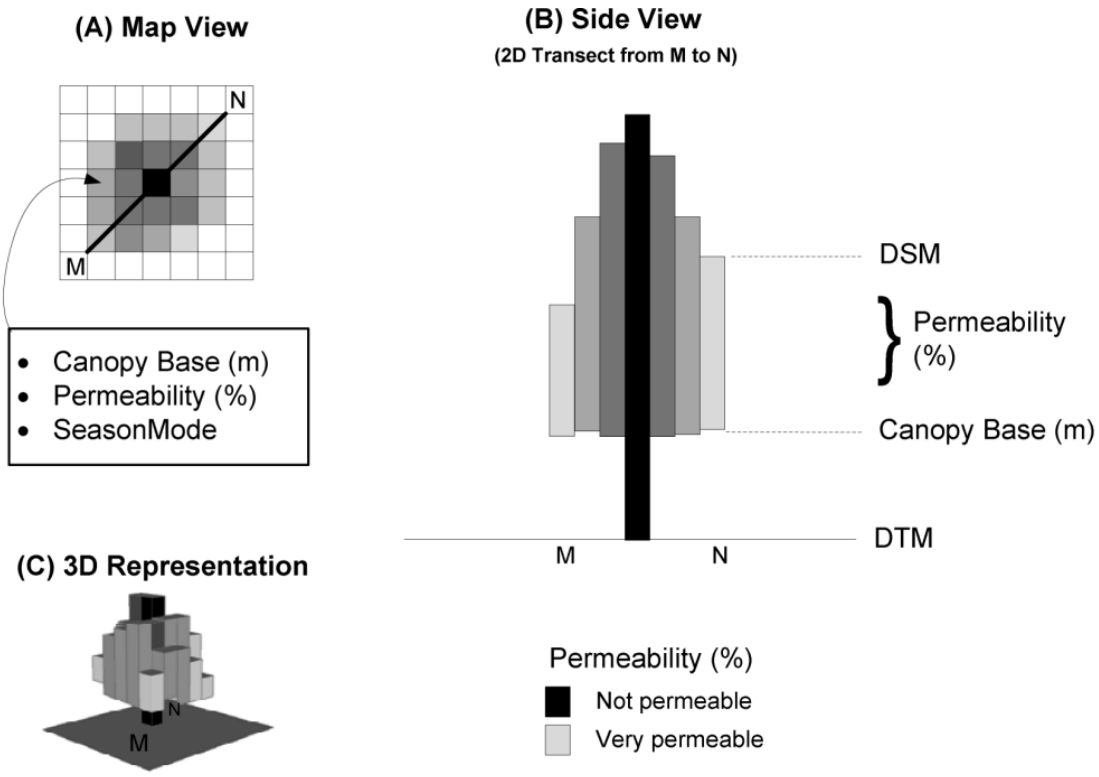


Figure 4.7: Raster Vegetation Information Representation of Trees

a) Map view of raster permeability for cells with transect marked from M to N.
b) Side view of raster tree model through transect M to N. The DSM provides the height values for the tree canopy, and pixel attributes are used to store the permeability and canopy base height above DTM.
c) 3D rendering of the raster tree.

4.4.3 Modification to the Line of Sight model to accommodate Vegetation Information

To accommodate the new information layers a new visibility model is required. The model projects the vegetated zones found between the observer and target onto the FOI, so that the façade area visible may be adjusted according to the coverage index, giving an estimate of the area of visible regions on the FOI (Figure 4.8). The FOI is considered as a set of target columns at the resolution of the raster DSM, in this case 1 metre resolution. The visibility is calculated for each target column determining the area clearly visible, and that behind vegetation as a function of the coverage index. As an example if the coverage index for a ray at the FOI is 50% and vegetation covers half of a target column, then the final area visible for that column would be 75% of the column height (50% of 50% in vegetation zone + 50% under and above the canopy).

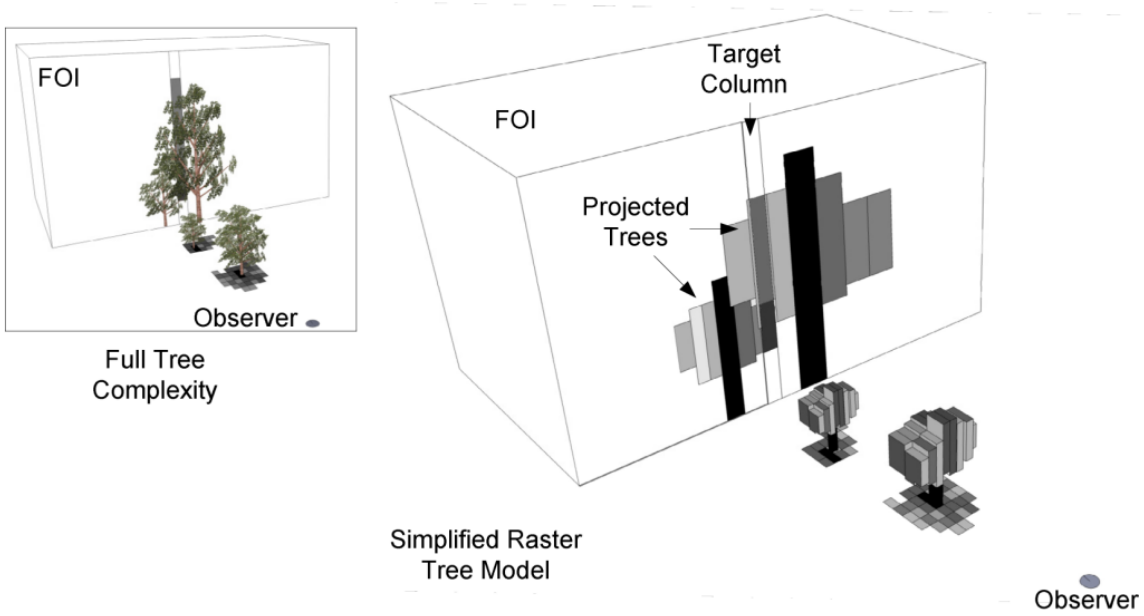


Figure 4.8: Projection of Vegetation on to the FOI Façade

The method to calculate the vegetation impact is as follows (supporting Figure 4.9):

a) The angle to the steepest non-vegetated DSM cell is projected on to the FOI to find the Not Seen (NS) zone.
b) The steepest gradient to a vegetated cell between the observer and target is determined, and the lowest canopy angle. This defines the Above Canopy (AC) and Under Canopy (UC) zones. The rays for these regions are subjected to an atmospheric rate of decrease. The NS zone always takes precedence when the other zones are defined.
c) The region on the FOI between the base of the AC zone and the top the UC zone is defined as the Through Canopy (TC) zone.
d) The vegetated regions along the path between observer and target are recorded (e.g. A,B,C) and the intercepts at the FOI are calculated for the top (lines At,Bt,Ct) and base (lines Ab,Bb,Cb) of each vegetated region, defining the TC subzones. Vegetated regions whose projection falls completely within the NS zone are removed, and any base lines which fall in the NS zone are redefined using the angle to the top of the NS zone.

e) The vegetated zones are considered in order, furthest from the observer first. To simplify the process an average ray path is used to describe the impact of each vegetated region in each TC subzone (e.g. lines Ax and Ay in detailed view). The permeability values from each cell are applied to the ray to calculate its coverage index, by multiplying the current coverage index by the permeability value, such that a ray experiences an exponential decay. The formula may be written as Coverage Index = $100(p_1)(p_2)...(p_n)$, where p is the decimal permeability rate in the range 0 to 1 which determines the likelihood of the ray passing through that cell (see Figure 4.6 for more details).

f) The TC zones are assigned an appropriate coverage index value which corresponds to the presence of vegetation along the LoS. For example the zone between Ct and Cb is assigned the coverage index calculated from ray Cx, the region between Bt and Bb is assigned by Bx, while Ax and Ay are used to assign the remaining A zones. Any regions not assigned a value are known as Inter-Canopy (IC) zones and treated as visible through clear atmosphere (same as AC and UC).

g) The coverage indexes (Ax,Ay,Bx,Cx) are applied to the column zone areas and summed to give a single façade area visible value for the TC zone.

h) Any rays which encounter trunk zones will be terminated.

i) Any rays which have a coverage index of below 1% are terminated and the zone is considered as not visible. As rays fade exponentially this is useful for improving performance for rays which are tending towards 0%.

j) The process is repeated for all target columns within an FOI to give a single visible façade area. A perceived area is also calculated which considers the observer‟s viewing angle and distance (see Section 4.3.2). For interior target cells within a FOI (i.e. not on the FOI boundary) a large proportion of the zoning will be NS, as the majority of the column is blocked by exterior cells.

The pseudo-code equivalent for this model is shown next.

Method CalculateVisibleArea_IncVegetation()

```
# Calculate the Zones on the Target which are:-
# Above Canopy - AC
# Through Canopy - TC
# Under Canopy - UC
# Not Seen - NS

TarZ = Target_DSM

#Calculate the Not Seen (NS) Region based on Close Horizon
A1= FindSteepestAngle (Non-Veg Cell)
NS = ProjecttoTarget (A1)
IF (NS > TarZ)
  RETURN 0 #target not visible so return to main routine

#Calculate the Above Canopy (AC) Region based on Vegetation Tops
A2 = FindSteepestAngle (Veg Cell)
TCTop = ProjecttoTarget (A2)
IF (TCTop < TarZ)
  AC =  TarZ - TCTop
ELSE
  AC = 0

#Calculate the Through Canopy Zone Sections - use the LOS from Observer to Target
#noting each vegetation block en route from these calculate the intercepts with the target

FOR EACH VegBlock in LOS (Observer, Target)
 Vt = FindSteepestAngle(VegBlock)
 Vb = FindLowestCanopy (VegBlock)
 TCList.Add(ProjecttoTarget (Vt))
 TCList.Add(ProjecttoTarget (Vb))
END FOR LOOP

#Sort TC Zones by Elevation
TCListSorted = TCList.Sort()

#Calculate the Avg Ray Coverage Per TC Zone
FOR EACH InterceptZ in TCListSorted(0,TCListSorted.Length-1)

 IF InterceptZ > NS
  #GET MidPt of Each TC Zone
  InterceptZ_next = TCListSorted(TCListSorted.Length+1)
  ZoneHeight = (InterceptZ_next - InterceptZ)
  MidPt = ZoneHeight / 2

  #calc ray coverage along path to target at designated MidPtZ height (raycasting)
  #if coverage below 1% then stop search and return value of 0
  RayCov=CalcRayCoverage(MidPtZ, 1)
  RayCoverageList.Add(RayCov,ZoneHeight)
 END IF

END FOR LOOP

#Total Target Area Visible
FacadeAreaVisible = AC

#TC, UC, IC zones
FOR EACH Zone in RayCoverageList
  FacadeAreaVisible += RayCov * ZoneHeight
END FOR LOOP

RETURN FacadeAreaVisible
```

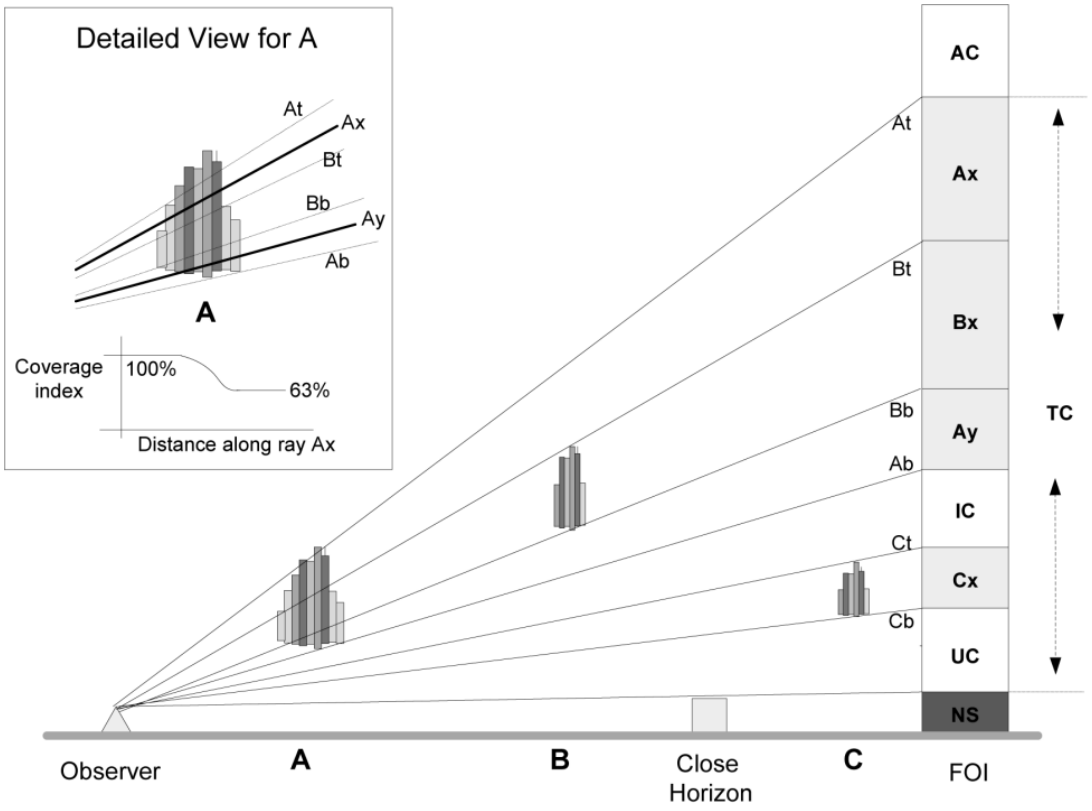


Figure 4.9: Line of Sight Model – projected vegetation zones (A,B,C) onto FOI

To describe the vegetation impact across the FOI façade a new visibility metric is introduced, which divides the FOI into thirds. The zone definitions are created without considering the visibility of the structure, but instead use the extreme FOI target points which create the widest angle from the current location. This is so the visibility results, which are summarised according to these zones, may include the percentages for area visible, visibility through vegetation, and hidden from view, as shown in Figure 4.10. The output quantifies the extent and pattern to which an FOI is blocked from the current view and may be used by LBSs to build natural egocentric descriptions, such as "the gallery is visible in front of you, although partially hidden by vegetation on the right side".

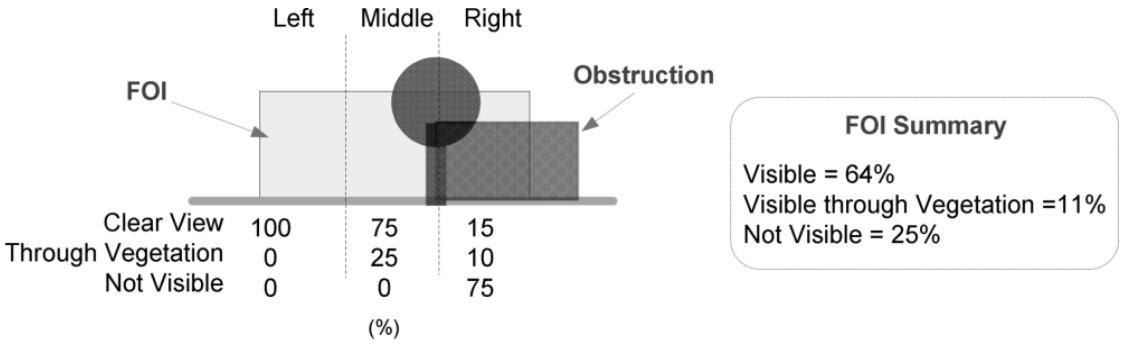


Figure 4.10: Summarising Visibility across an FOI in Thirds by Viewing Angle

The next section demonstrates the use of the modified visibility model in a vegetated urban area through a number of trials which illustrate how the adaptations improve the modelled results, and are able to accommodate seasonal variation, views under bridges, and also a demonstration of how the new metric may be used to describe a scene.

4.5 Examples of Visibility Modelling with Vegetation

4.5.1 Differences between Visibility modelling with and without vegetation map

The first example demonstrates the difference in the visible region when the observation location is sited under vegetation. The visibility of nearby objects can be massively over or under reported when only a DSM is used, as depicted earlier in Figure 4.2. Typically views are either extended as if the user was positioned at tree top height, or very limited by nearby undulating canopy. Figure 4.11 shows an example where the horizon is drawn for an observer standing under a tree. In the case without the vegetation model the results are irregular, displaying almost no view at the first location and a wide view for a point 1 metre further north. In reality the views are almost identical as shown by the model which includes the vegetation information. The model which includes the vegetation information is therefore more stable, and similar to real world experiences.

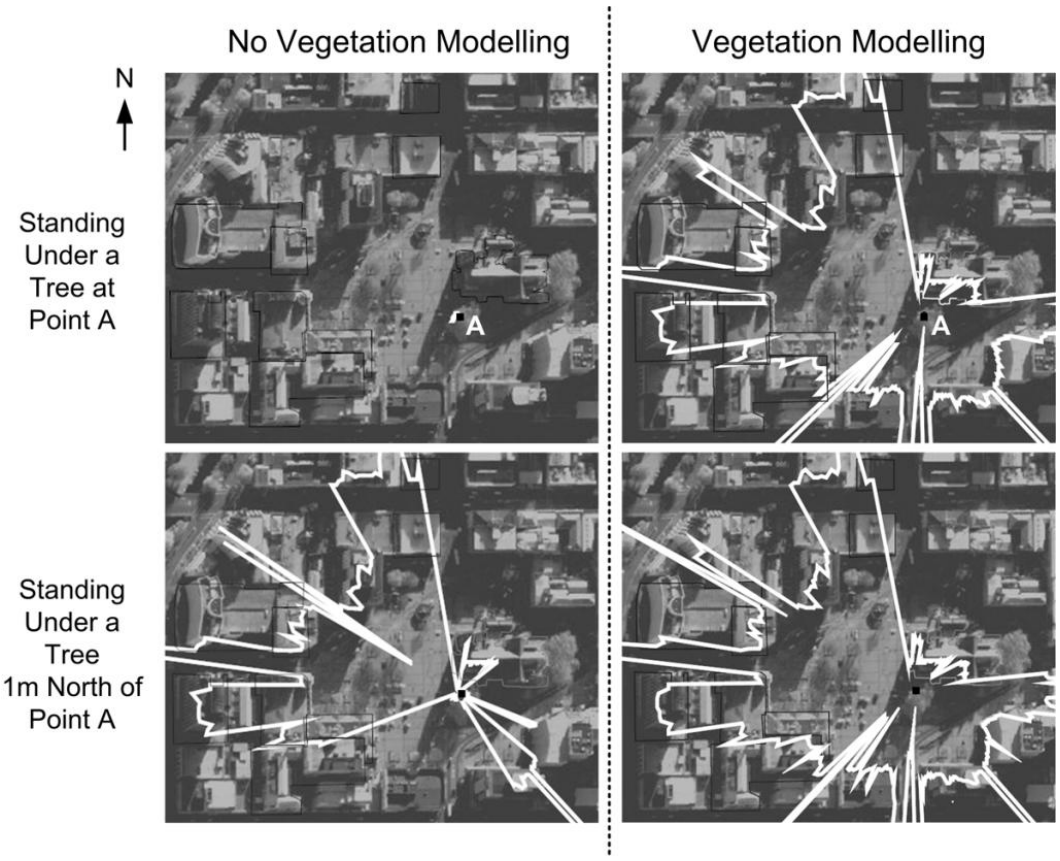


Figure 4.11: The Horizon as Modelled with and without a Vegetation Map, from a Location under a Tree

The new model also permits calculation of visibility under canopy cover. The visibility of each target is divided into component parts to describe how much is clearly visible and visible through vegetation. LoSs passing under the canopy are considered as clearly visible, and the *coverage index* decays at the atmospheric rate. Figure 4.12 shows the modelled visibility of a landmark at three locations, the latter two both including views under the canopy layer. The *Façade Area Visible* metric is an indication of how much of the building is visible, while the *Perceived Area* metric factors in viewing angle and distance. Prior to the inclusion of the vegetation map the FOI was not considered visible from the vegetated area (Figure 4.12(iii)).

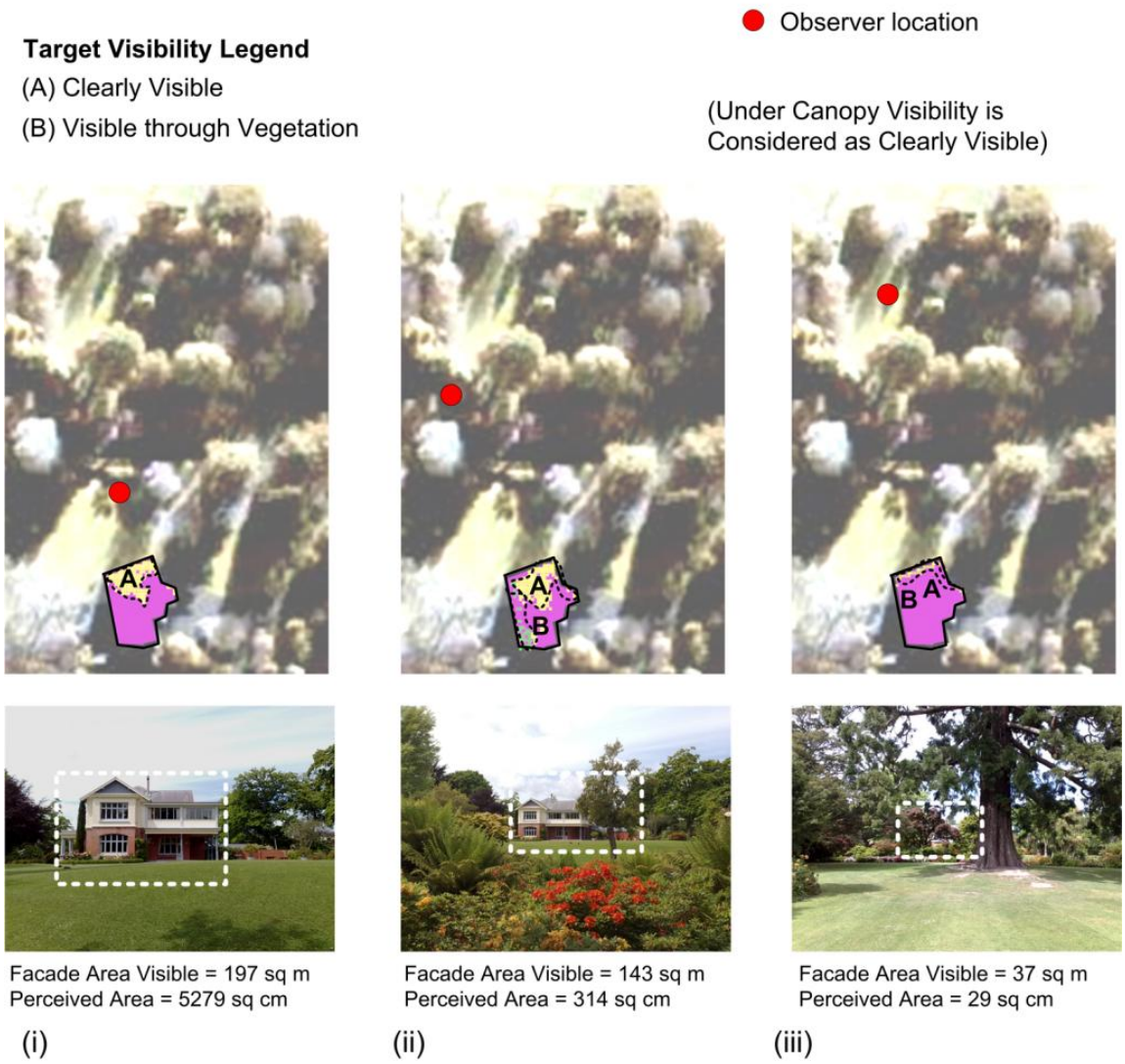
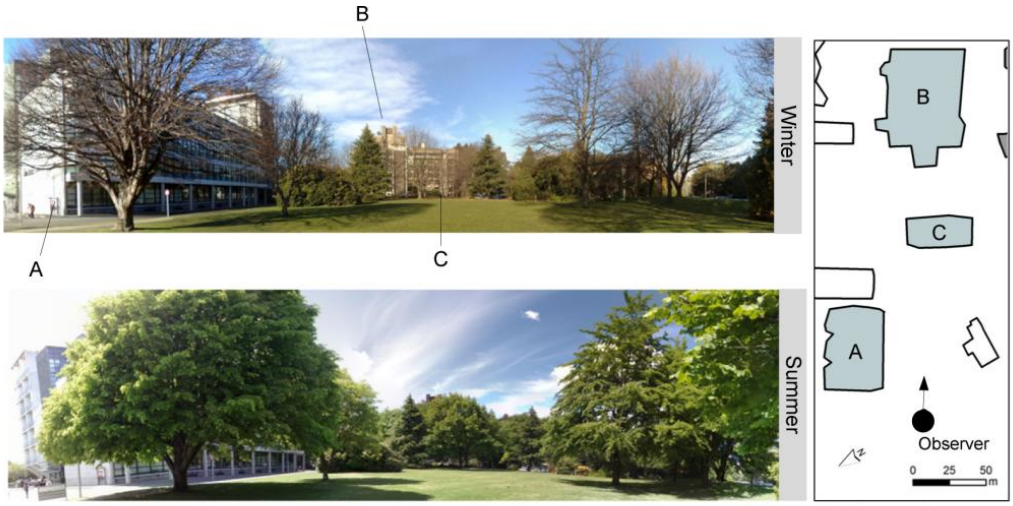


Figure 4.12: Modelling the Visibility of a Landmark from Three Distances

4.5.2 Taking Account of Seasonal Vegetation Changes

As outlined in Section 4.4.2 the model accommodates a simplified approach to cater for seasonal vegetation changes. Each cell in the vegetation map is encoded with a Boolean value to reflect if it responds to change between seasons. Under canopy results are not affected by seasonal variation, nor are Evergreens and tree trunks which exhibit the same visual properties throughout the year. The cells which are affected by season are automatically assigned a standard single minimal visual permeability rate for the winter mode, allowing the rays to pass through to a greater extent, simulating the lack of winter foliage.

This is a highly simplified approach to seasonal change, as trees drop leaves at different times of the year and to varying degrees. Although the model could be easily modified to store a seasonal change date, and fade between winter and summer permeability values, it is not yet practical to collect such data for large city areas. Therefore this example only models a basic switch between the extreme summer and winter foliage densities.



	Area (sq m)	Perceived Area Total (sq cm)	Clear (sq cm)	Through Vegetation (sq cm)
Winter				
A	1025	3421	2276	1145
B	257	57	57	0
C	810	538	242	296
Summer				
A	819	2568	2276	292
B	257	57	57	0
C	386	252	242	10

Figure 4.13: Comparing Summer and Winter Modelling Results

The modelling of visibility for 3 FOIs was carried out in both winter and summertime, as shown in Figure 4.13. The *Area* column describes the total building façade visible, irrespective of viewing distance or viewing angle. The *Perceived Area Total* column is calculated by factoring the *Area* visible with the viewing distance, viewing angle, and LoS ray *coverage index*. A more detailed explanation of the differences between metrics may be found in (Bartie *et al.* 2010). The *Clear* and *Through Vegetation* columns show the division of *Perceived Area* into the component parts, for those rays passing through vegetation and those only decaying at atmospheric rates. A number of observations can be made from the results. The *Perceived Area Visible Through Vegetation* increases dramatically in winter time as expected with minimal foliage to block the LoS rays which pass through the vegetated zones. Building B experiences no change between summer and winter as the only parts visible from this location are behind Building C and above the tree line. The perceived *Clear Area* does not change between summertime and wintertime for any of the buildings, as the view above and underneath the canopy are not affected by the seasons. Building A shows less change between seasons than Building C, because the northwest wall (left side of photograph) is clearly visible all year around, and not hidden behind vegetation. Building C experiences the most dramatic change between summer and winter, with the largest increase in area visible through vegetation.

It is also noticeable that the results for Building C tend to underreport the impact of vegetation. As the photograph in Figure 4.13 shows, Building C was mostly hidden from view in

summertime with only a small proportion of the building visible under canopy. The calculated area difference between summer and winter indicates a doubling in the visible building area, while the photographic evidence suggests the change should be greater. Upon further investigation it appears that the LiDAR dataset failed to capture the complete canopy shape for the large tree in front of Building C. Full waveform LiDAR (Reitberger *et al.* 2006) and improved data capture technologies will assist in the future, as more detailed tree profiles will be available to the model.

4.5.3 Modelling Visibility under Bridges, and Overpasses

Aerial LiDAR surveys only capture information in the line of sight, therefore clearance heights under bridges and overpasses are not available. However with minor coding alterations a special case may be included into the visibility model to accommodate LoS passing under such objects. To do this the tree trunk model was adapted such that the canopy height stores the value of clearance above the DTM, allowing the LoS to pass under the DSM elevation at the atmospheric decay rate. The permeability class is used to denote the cell as a special case so the model does not consider the LoSs passing through these cells as having encountered vegetation. An example of this is shown in Figure 4.14, whereby the majority of the building would have previously been considered out of sight.

The procedure to update a map region is fairly simple, requiring only the cells under marked bridges and walkways to be assigned the special permeability class. The clearance heights can be set to a generic value for vehicles or pedestrians appropriately, or refined through a ground survey.

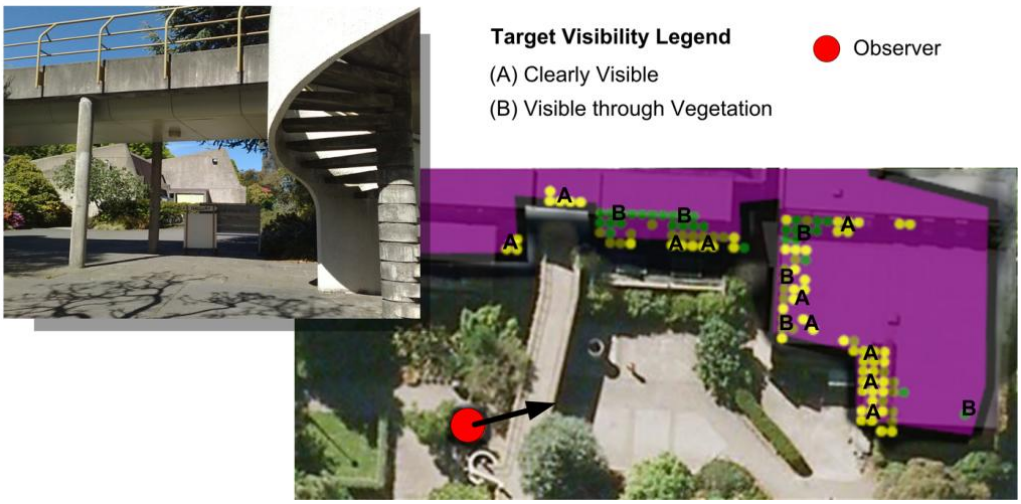


Figure 4.14: Adapting the vegetation model to cater for Underpasses and Bridges (Background imagery sourced from Google Earth™)

4.5.4 Reporting the Vegetation Distribution across a Feature

The vegetation distribution in front of a FOI may be described by dividing the maximum field of view for the FOI, regardless of its current visibility, into thirds. The interception heights (Figure 4.9) are used to summarise the percentage clearly visible and that visible through vegetation, into the three zones. The ray *coverage index* is not applied to the result, as the metric is intended to provide a quantitative description of the vegetation's impact on the view across the FOI façade from a given location irrespective of vegetation type or density. Figure 4.15 illustrates this with an example where the right side of the FOI is mostly hidden by vegetation, the central third is slightly hidden behind vegetation, and the left side is totally clear. Any views under the vegetation canopy are classed as clearly visible.

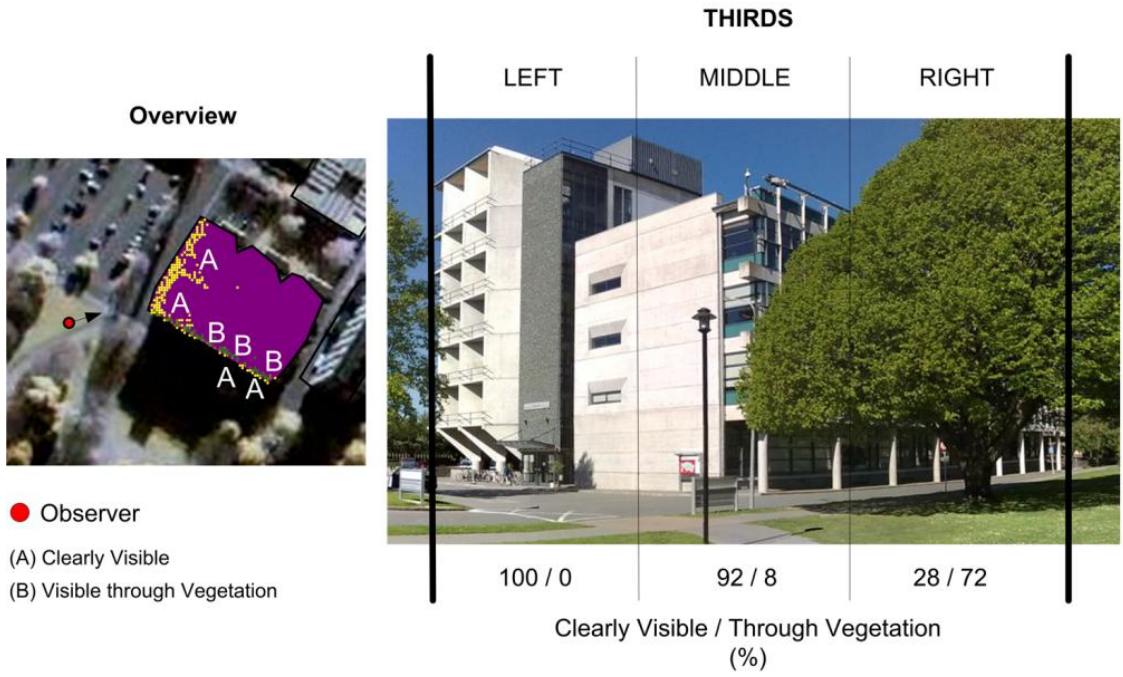


Figure 4.15: Quantifying the Vegetation Distribution in front of a Feature

4.6 Conclusions and Future Work

The introduction of LiDAR sourced DSMs has permitted urban visibility modelling to move from isovist (Tandy 1967, Benedikt 1979) to viewshed methods (Tandy 1967, Lynch 1976), permitting the visual exposure model (Llobera 2003) to be applied in urban environments. However the model is subject to errors which result from the representation of vegetation in LiDAR sourced DSMs, and therefore a number of modifications are required so that LoSs may be modelled through vegetated zones. While vegetation has been considered in visibility models before (Dean 1997, Llobera 2007a, Liu *et al.* 2008) consideration has not been given previously to how the visual exposure model may be adapted to accommodate supplementary information on vegetation density and distribution.

This chapter implements an urban visual exposure model with a number of modifications and novelties. The impact of vegetation on the view of landmarks is incorporated by considering raster tree models, which provide canopy base height, tree density, and seasonal variation information on a cell by cell basis. In addition both a DSM and DTM are used such that the observer may always be placed at ground level, and not incorrectly at tree top height when walking underneath trees. Furthermore the LoS is permitted to pass partially through, and clearly under, the canopy layer. A new visual exposure metric is proposed and demonstrated which describes the distribution of vegetation in front of FOIs from an egocentric viewpoint. This is useful for building narrative descriptions for a feature, and may be used by LBSs to describe scenes to a user, or identify targets by the presence of foreground vegetation.

The model has a number of limitations associated with collection and storage of the permeability values. The image based thresholding technique is time consuming, and defines average permeability values for vegetated regions possibly missing significant density variations. The model is also limited to a single set of attributes which describes the canopy vertically above each cell. This could be improved by defining many attribute sets over different vertical ranges, or using voxel data structures, thereby accommodating multiple canopy layers and improving the fidelity of the model. However the biggest difficulty is in collecting the detailed tree foliage densities, trunk and branch locations. Ground based LiDAR surveys for city regions could be used to capture more accurate tree models (Omasa *et al.* 2008), and it may also be possible to distinguish different tree types in urban settings using full waveform LiDAR (Reitberger *et al.* 2006) and thereby automatically determine deciduous trees.

The model may be run in any urban environment where the spatial information layers are available, however there is a considerable computational impact on the simulation performance when introducing vegetation into the visual exposure model. Previously many LoSs were terminated prior to reaching their designated targets, as they encounter visual blockades where trees exist in the DSM. As a result of the supplied vegetation layer these rays are now modelled through the vegetation to their designated targets, dramatically increasing the number of computation steps required per ray. Furthermore the number of lookups is tripled from a single DSM to the DSM, DTM and vegetation layer adding computational overhead. For near real-time applications the model may be parallelised across multiple processors (Teng 1993, De Floriani *et al.* 1994b), or the results cached for rapid retrieval.

It is rather difficult to empirically test the model's output quantitatively, but this should form the basis of future work, perhaps by conducting user trials which compare the model's output with observers' opinions.

4.7 Acknowledgments:

The authors wish to thank Matt Faulk and anonymous reviewers for comments on the chapter, and the Christchurch City Council for use of their LiDAR dataset. The research was supported by funding from the University of Canterbury, New Zealand.

Chapter 5: Route Ahead Visibility Mapping: A method to model how far ahead a motorist may view a designated route

5.1 Summary

We examine how visibility analysis may be used to calculate the extent of the route ahead visible to a motorist. To do this a Digital Surface Model sourced from a LiDAR dataset is used, which includes building and vegetation profiles as well as topography. Once the route has been designated the visibility along the path is modelled, in the direction of travel, from the driver's viewpoint. The visibility analysis considers all surrounding objects during the calculation, but reports only how far ahead the route may be viewed, and not the total landscape area visible. The map accompanying the article shows the results from using this method for a route in Christchurch, New Zealand, from which we identify a number of interesting results along the route. This method may be used such that navigational devices customise announcements to the most opportune times when the driver's workload is minimal, and decision points are in view rather than at predefined distances. It could also be used to notify the driver of the safest overtaking locations or those areas of limited visibility.

5.2 Introduction

In this research we look at how visibility analysis may be used to provide information on how far a motorist can view the route ahead. This technique can be applied to a number of areas including the planning of the safest routes for high speed emergency response, or to identify the safest road sections in a journey for an overtaking manoeuvre. Navigation systems may also use this to report when navigation decision points come into view, or to give advanced notice of restricted visibility junctions. Here we describe the method and give an example of its use in an urban area, where road networks are more dense and complex than in rural regions. We map the visibility results to show the outcome of the analysis.

Visibility studies within GIS require access to a digital elevation dataset for the area of interest. These are considered to be 2.5 dimensional, recording a single elevation value (z) for each location (x,y). Depending on the internal structure of the dataset, this may be called a Digital Elevation Model (DEM) if gridded or a Triangulated Irregular Network (TIN) if irregular (Kumler 1994). In this work we use gridded datasets and we make a distinction between Digital Terrain Models that represent the "bare earth" surface, and Digital Surface Models (DSMs) that include trees, buildings, and other manmade features.

There have been numerous previous studies on urban visibility, these have tended to use the two dimensional boundaries of buildings to calculate isovists (Tandy 1967, Benedikt 1979, Turner *et al.* 2001). To more closely model urban visibility it is necessary to use a DSM at high resolution, such that building and vegetation profiles are captured accurately. Light Detection and Ranging (LiDAR) remote data capture techniques have been shown to be suitable for this in urban studies (Palmer and Shan 2002, Rottensteiner and Briese 2002, Bartie and Mackaness 2006).

In previous related work visibility has been used as a criteria to calculate the most scenic, or hidden routes (Lee and Stucky 1998, Stucky 1998). It has also been used to find the most salient landmarks at decision points, those which can be seen from greater distances on the approach to junctions (Winter 2003). Computer vision techniques have been used to estimate the distances to other road users in real time (Betke and Nguyen 1998, Leibe *et al.* 2007), and measure visibility in foggy (Hautiere *et al.* 2007), or snowy (Matsuzawa *et al.* 2009) conditions. Signage clarity and the benefits of icons have been measured(Kline *et al.* 1990, Kline and Fuchs 1993), as well as the visibility of road markings (Zwahlen and Schnell 1999).

This research focuses on modelling the visibility of the immediate road ahead for planned routes. The results do not represent the visibility of the surrounding landscape, but reflect the clarity of the next road section of the prescribed journey. The route itself may be manually entered, or calculated using shortest path algorithms and include highways, lanes, bridges and so on. Currently modelling visibility under bridges, or through tunnels, is not possible using LiDAR sourced DSMs, as elevation data is not captured for these occluded regions. However supplementary information layers may be used to enhance the surface model to remove these restrictions in the future.

The visibility results reflect the road layout (i.e. width, corners and junctions), as well as the topography and surrounding built form. There are a number of applications for this, including navigation devices which are able to report when a navigation decision point has come into view rather than base the announcement on predefined distances. It also provides knowledge which can assist in identifying safer routes with higher overall advanced route visibility. Furthermore it can provide a driver with advice of the upcoming overtaking opportunities, and highlight danger areas with limited visibility.

5.3 Preparing the Digital Surface Model

LiDAR datasets can be used to produce high resolution surface models, which capture ground topography, vegetation, building profiles, and surface objects such as cars and buses. In our calculations we wish to use the surrounding surface model to restrict the visibility for points along the route; however we wish to calculate this based on a clear road without the vehicle

elevations present at the time of LiDAR data capture. It was therefore necessary to process the DSM to remove vehicles from the roads, while attempting to maintain traffic islands, roundabouts, overpasses, and vegetation barriers. To do this we set a road boundary analysis mask using a vector road polygon boundary layer, and calculated the slope angle from the DSM. The slope angle reveals the rate of elevation change between cells; those with a very different elevation to their neighbours are highlighted as having a high slope angle. In our case as roads are designed to be fairly smooth the high slope angles denoted surface objects such as vehicles and elevated roundabouts. By using a filter to remove the lower slope angles from the dataset we can ensure that any hump-backed bridges and hills were not included as road surface objects layer. As this method only identifies the boundary of surface objects it was necessary to use a fill process to infill the centres of the designated regions, resulting in a layer denoting road surface items consisting of both vehicles, and road vegetation barriers. Using this layer we could mask out those cells in the DSM roads which were attributed to non-road surface items replacing them with No Data values. These No Data values could then be replaced with a focal mean value from the surrounding region. The result is a smooth road surface elevation value which maintains topography without the local distortions due to road surface items.

To restore the road vegetation barriers and vegetated roundabouts we then used a Normalised Differential Vegetation Index (NDVI) layer, created from a Quickbird dataset. This satellite captures 2.5 metre resolution imagery in 4 bands including Near Infrared, and a panchromatic band at 60cm resolution for image enhancement. By using the NDVI layer we were able to locate the road cells which were obscured as a result of vegetation. A mask was created so that the original DSM values could be replaced in our smoothed road dataset for these vegetated cells within the road boundary definition. Finally, a new DSM was constructed from the original DSM non-road values and the new road surface values, resulting in a DSM for the entire city showing building profiles and topography without vehicles on the roads. This layer could now be used to calculate visibility along roads, while maintaining the visibility limitations imposed by buildings and vegetation. The process is summarised in Figure 5.1.

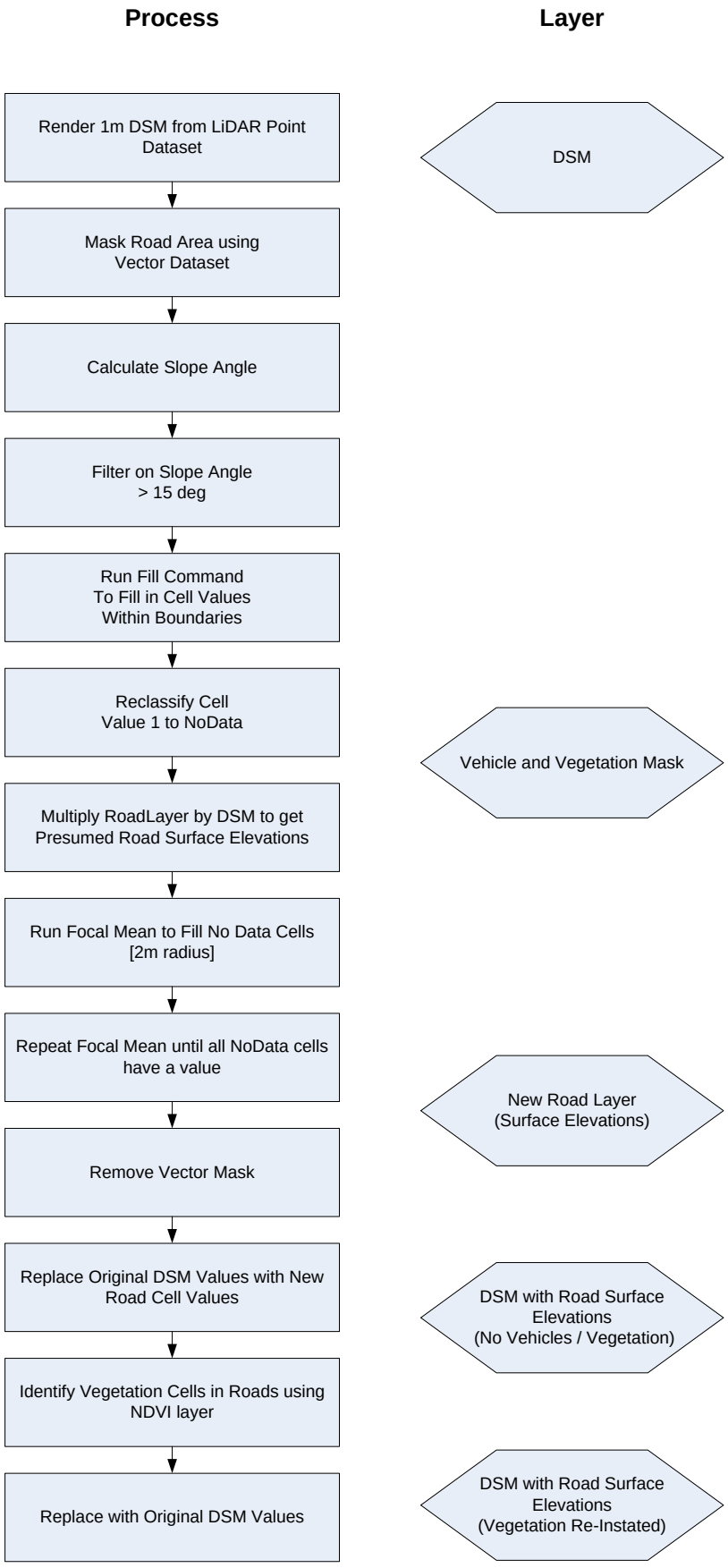


Figure 5.1: Process to produce Digital Surface Model without Vehicles on Road Surface

5.4 Visibility Analysis

Road visibility was modelled using a custom application based on an object based visibility modelling engine, able to calculate a number of visibility metrics at entity level (Bartie *et al.* 2008). In this case the road sections making up the designated route are set as the target objects. The observer and target locations are restricted to road locations, therefore the results are not an indication of the total area visible from each point, but limited to how much of the route ahead is visible. In this research we placed observation and target points along the route every 2 metres, limited the maximum search extent to a route distance of 1 kilometre, set the driver eye height to 1.5m from the ground, and the target height to be just above ground level (1m) to overcome any minor imperfections in the road DSM dataset. The target height can be used to model the visibility of other road vehicles ahead if this is desired.

The software reports two values per observation point, these are the total number of points on the road ahead visible, and the number of continuous points visible immediately in front of the current location until the road is first obscured. As each point is spaced at 2 metre intervals along the route we can therefore calculate the route distance visible in front of the observation point. Figure 5.2 shows the vehicle at Point A, able to view all points on the road ahead until Point B. The view of the next road section is blocked by a building at the side of the road on a corner. However from Point C the road is visible again. The two values calculated here are the continuous route distance from A to B, and the total visible route distance from A to B and then from C onwards until the maximum visible extent is reached.

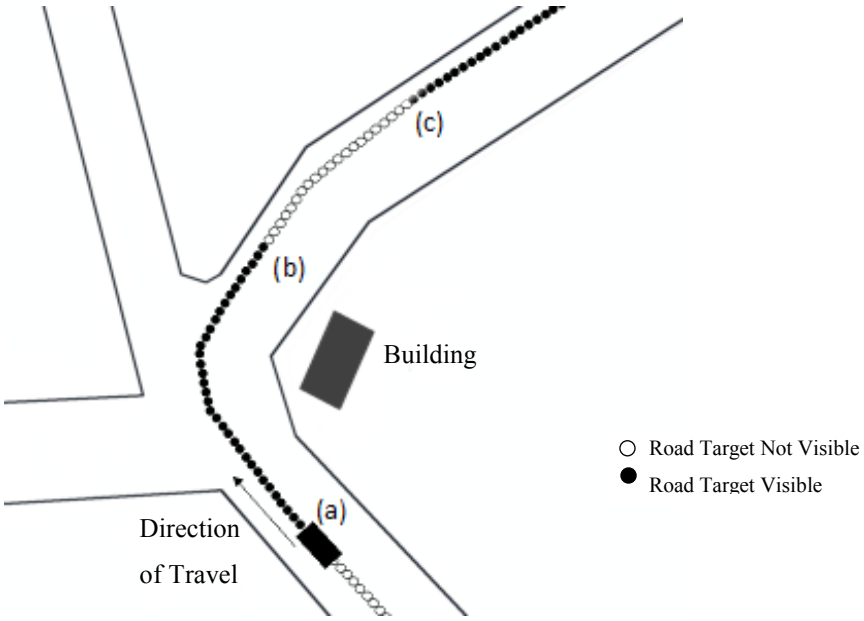


Figure 5.2: The Difference between Continuous Visibility and Total Visibility

In many cases the two statistical outputs are identical; however differences occur when the observer is only able to view partial sections of the road ahead due to undulating topography, or surrounding obstructions (e.g. buildings). An example of this is shown in Figure 5.3 where the Total road distance visible ahead is greater than the Continuous distance at Point A as depicted by track width. This is because the driver is able to see some of the forthcoming route across the garage forecourt. As a result the continuous road distance values are lower, reporting the visibility from the current location to the corner, while the total route visible value indicates that future sections of the route are visible beyond the garage. For identifying prime overtaking locations the continuous value would be most suitable, while the total value may be suited to applications which find the safest routes, or for cueing navigation announcements.

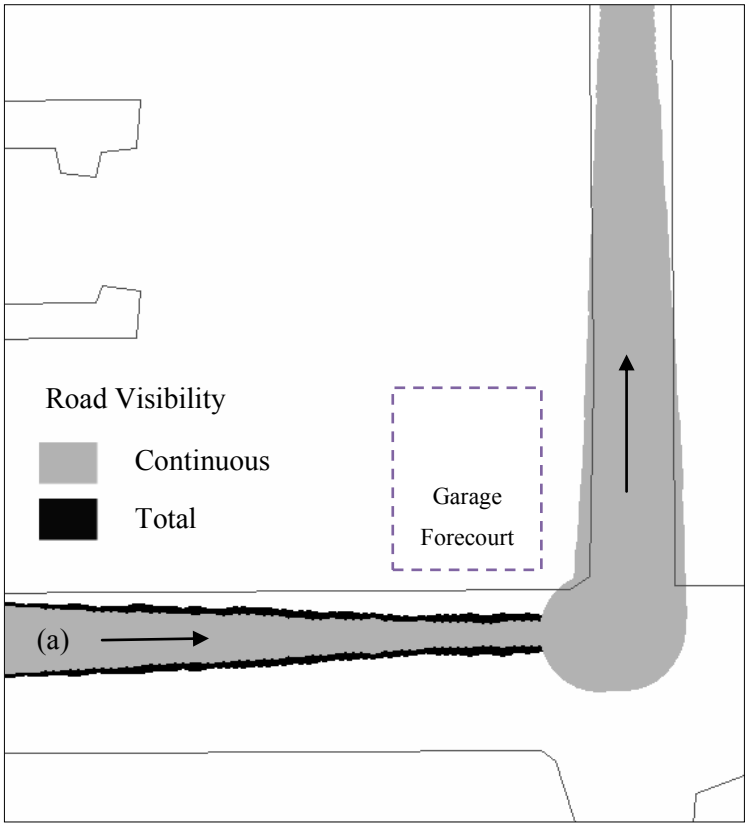


Figure 5.3: Example of Continuous and Total Visibility

In the following trial (Figure 5.4) the width of the line depicts the Total visibility of the route ahead. Note the gradual decrease in visibility towards junctions as would be expected, and the sudden increase during the navigation of corners. Also notice the drastic increase in visibility once the route has exited the city centre, moving onto a longer straight road section exiting on the west side of the map.

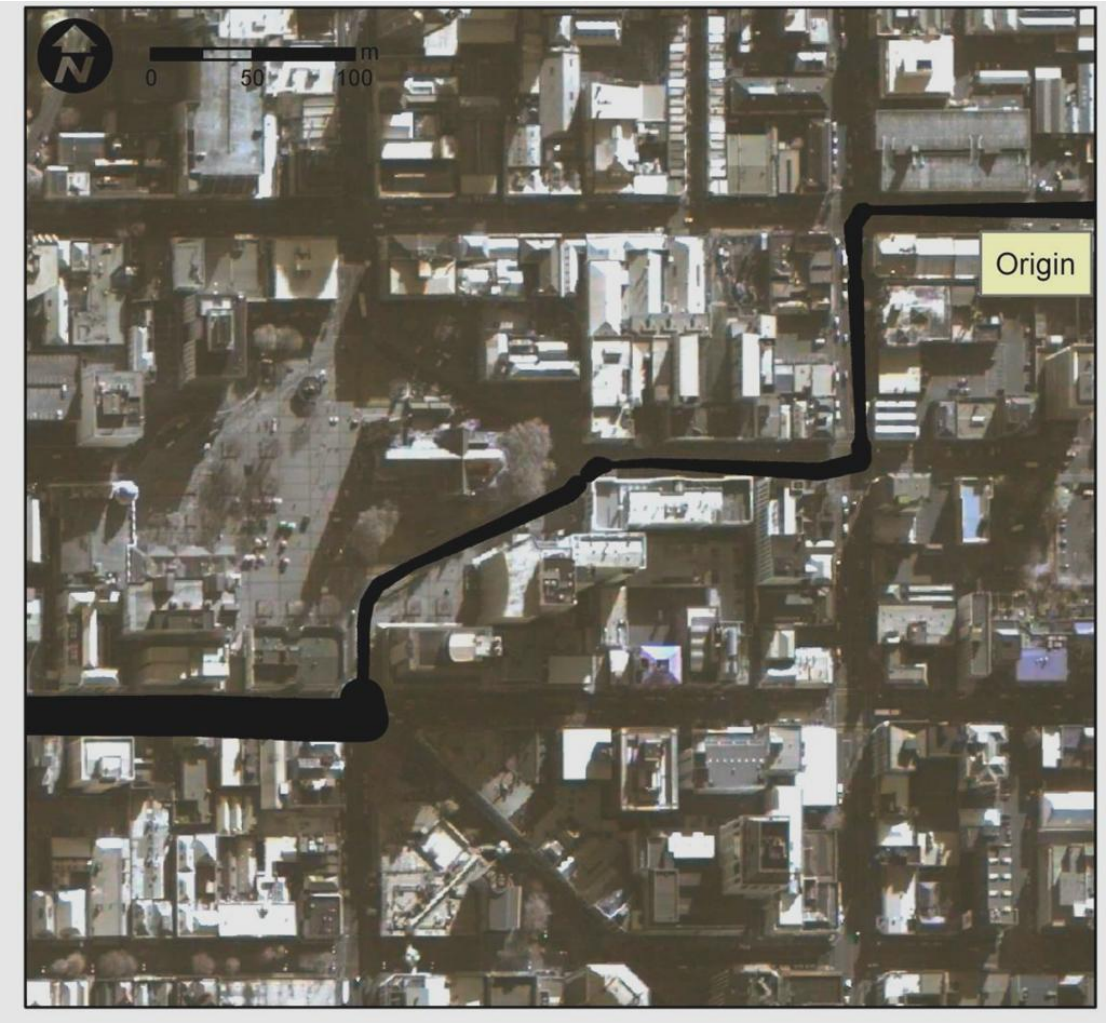


Figure 5.4: Example of Total Route Visibility in a City Centre

In the next example we model the visibility of the route ahead for a journey across the city of Christchurch, New Zealand. The journey encompasses a number of small bridges, an overpass, and several roundabouts. In the following section we review the results.

5.5 Results

Figure 5.5 shows an entire journey across the city of Christchurch, New Zealand. This map is also available as a high quality PDF download from the Journal of Maps website[1].

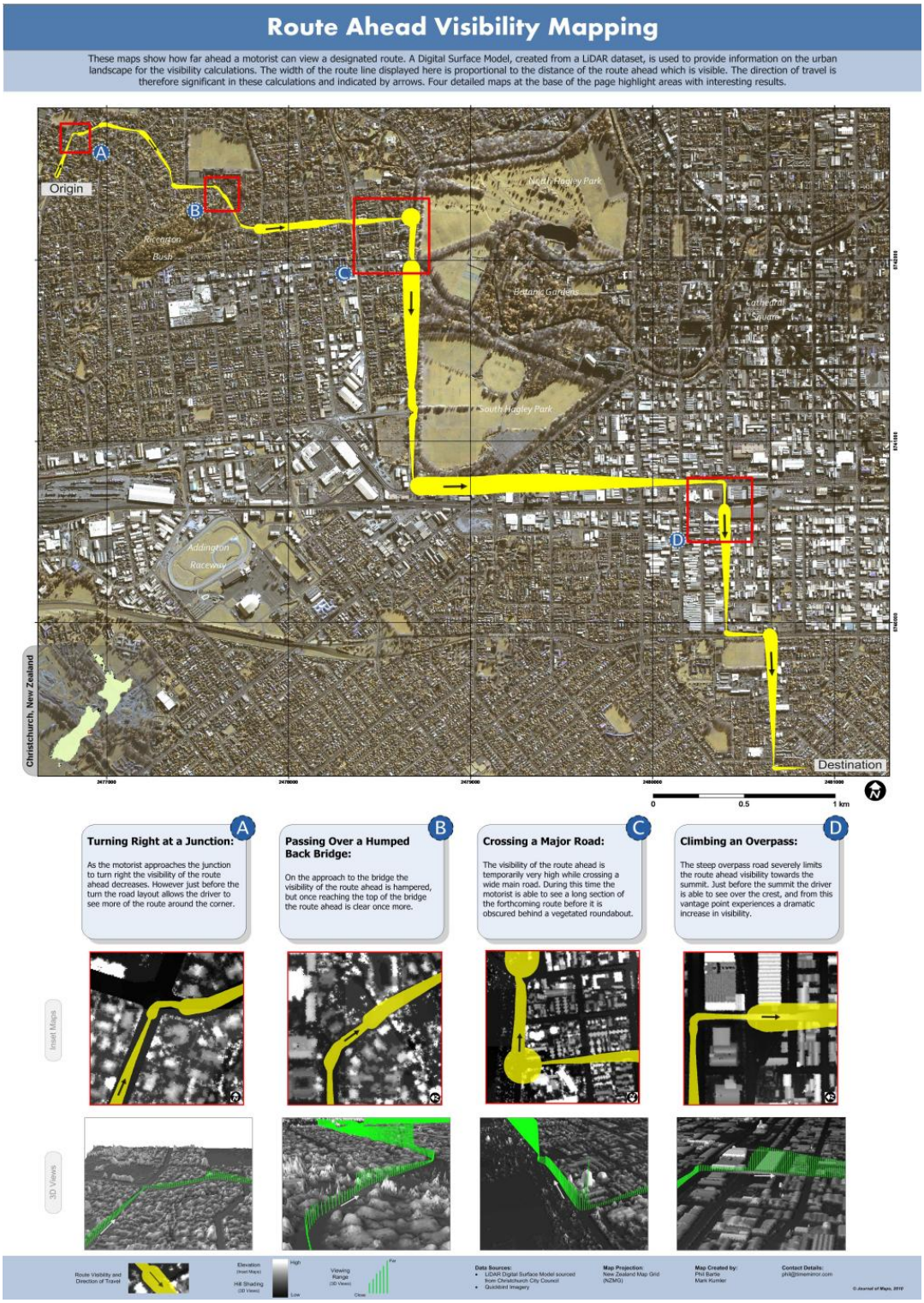


Figure 5.5: Route Ahead Visibility Mapping - Christchurch, New Zealand

[1] http://www.journalofmaps.com/viewMap.php?mid=1107

The total visible distance of the planned route ahead generally decays towards bends in the road, or junction points. However there are a few regions which require further investigation and explanation.

In Figure 5.6 near the start of the journey we turn right from the main road into a section of flat road with many corners. The visibility along this section is variable, determined by the tightness of the corner and road width.

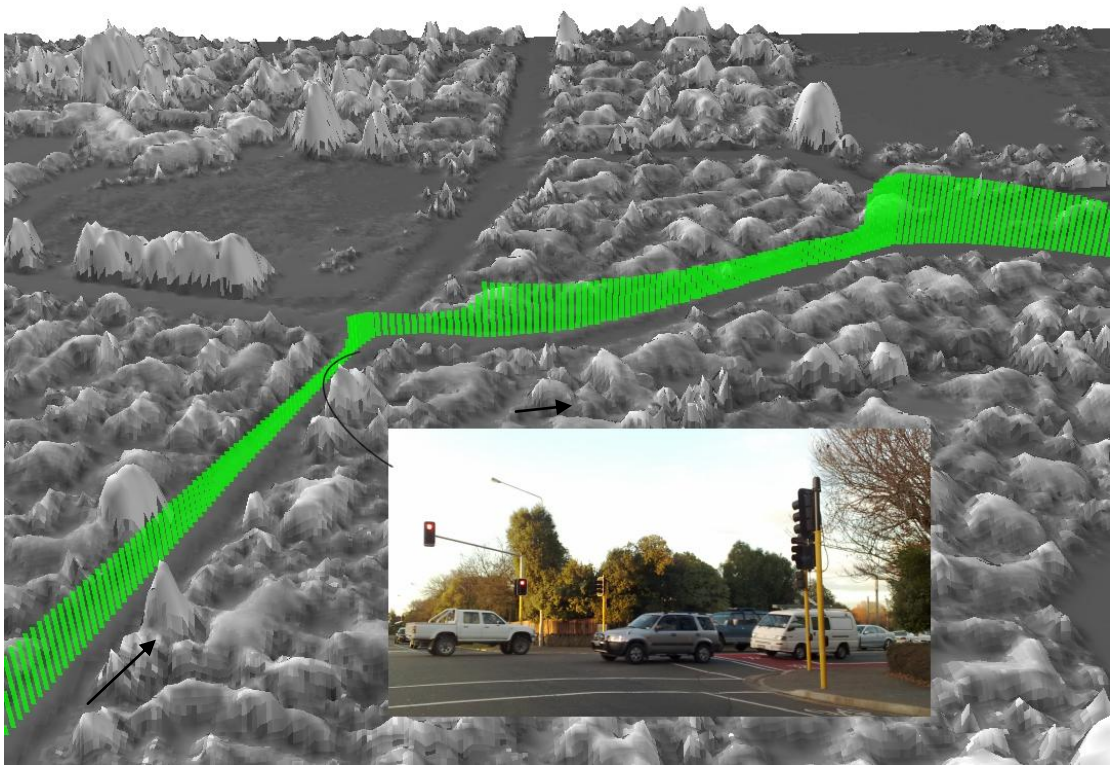


Figure 5.6: Visibility at a Junction

A little way further we pass over a short hump backed bridge which spans a small river. Figure 5.7 shows the dramatic decrease in visibility towards the summit of this bridge, and an increase in visibility just prior to crossing the summit when the driver's head would be above the crest of the hill.

Further on the path crosses a main street (Figure 5.8) and for a brief period a large extent of the remaining route extending towards (c) is visible. However once on the left side of the road the roundabouts at (b) and (c) limit the extent of the visibility of the route ahead and the route visibility drops. Once passed the roundabout (b) visibility increases rapidly, dropping towards the end of the road (c).

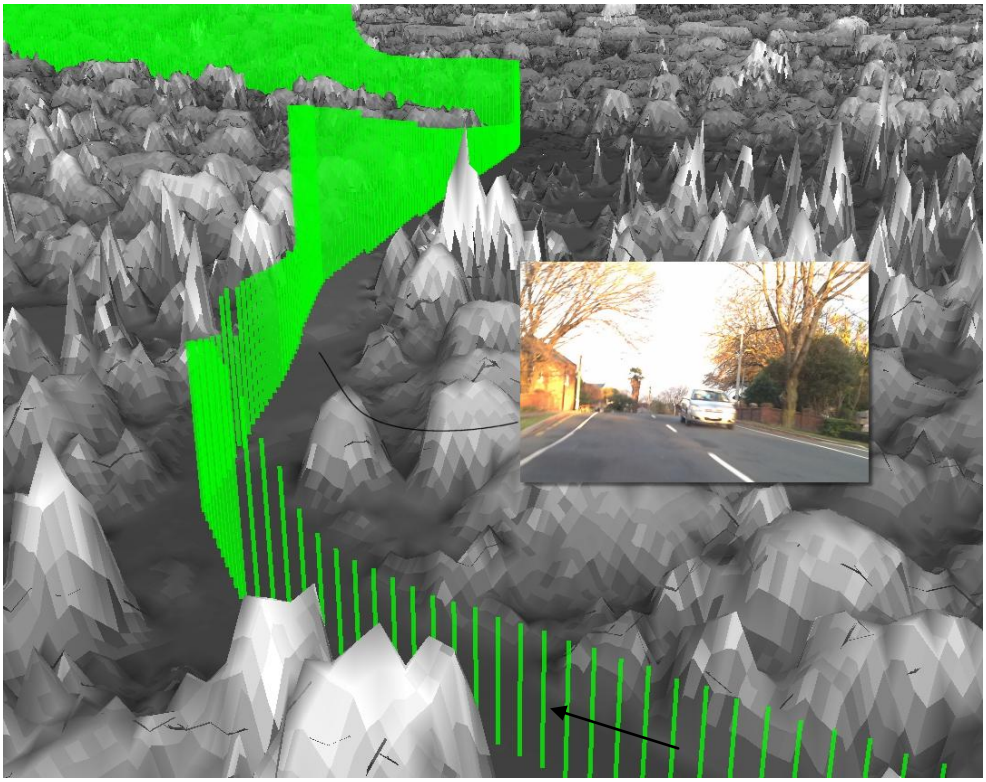


Figure 5.7: Visibility when Passing Over a Humped Back Bridge

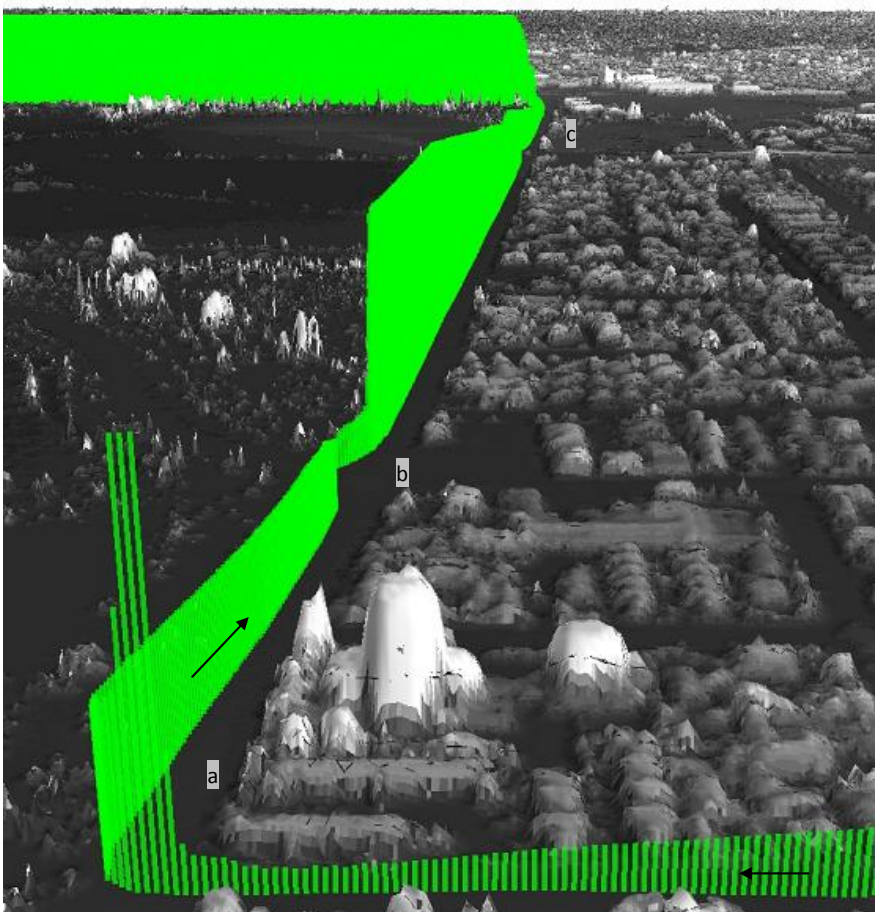


Figure 5.8: Visibility Turning Right Across a Main Road

In Figure 5.9 the view ahead gradually diminishes as the driver approaches turning point (a). After making the right turn the distance visible increases but diminishes again towards the brow of an overpass, point (b). Just before the summit the driver is able to see a greater extent of the road ahead clearly and the visibility results increase dramatically.

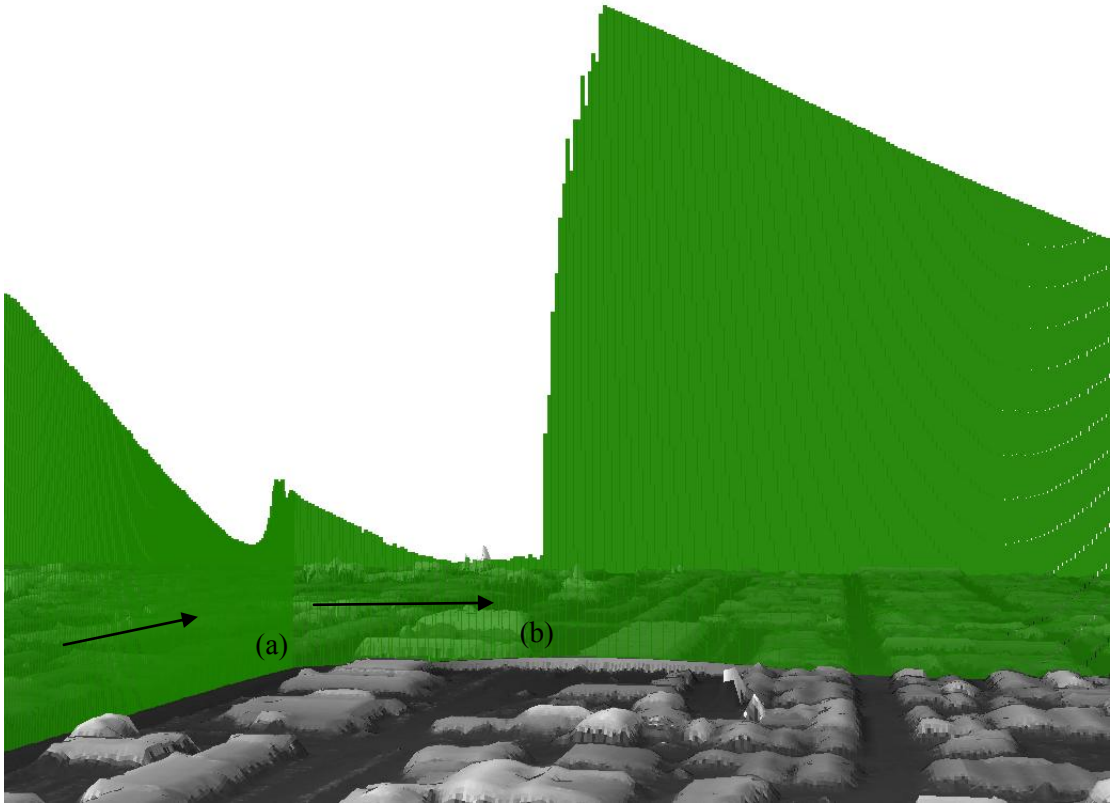


Figure 5.9: Visibility when Climbing an Overpass

5.6 Conclusion

The method outlined here is useful for quantifying how much of the planned route ahead may be seen from a given location. We have demonstrated this may be used for producing maps of route visibility, which could be used to highlight navigationally difficult areas. Here we are able to identify in advance those parts of the journey where the driver would need to travel more slowly due to limited forward visibility and those where the view of the oncoming planned route extends far into the distance.

The technique could be extended to the route planning phase, such that the quickest routes with highest route visibility could be calculated. This would be particularly useful for emergency services that travel at high speeds along city streets, and must consider not just their own ability to see ahead, but also the ability of other drivers to see the emergency vehicle.

Similarly, the technique could be incorporated into navigational devices to offer the driver information on the most suitable upcoming overtaking opportunities. Further research would be required to determine additional model inputs such as driving speed, road conditions, section

distances, and number of lanes. Such a system would be particularly useful for motorists looking to overtake in unfamiliar rural regions.

5.7 Acknowledgements

The authors wish to thank the Christchurch City Council for use of its LiDAR dataset, and the Geospatial Research Centre (www.grcnz.com) for research support.

Chapter 6: A Qualitative Model for Describing the Arrangement of Visible Cityscape Objects from an Egocentric Viewpoint

6.1 Summary

A projective spatial reasoning model is presented here, which is suitable for describing the relationships between the visible parts of buildings as seen by a street observer. It is intended for use in a Location Based Service (LBS) whereby users access geo-referenced digital datasets on location. Typically such applications filter data according to keywords and two-dimensional spatial reasoning, such as finding all hotels within 500 metres. However a LBS which in addition is able to reason from the user's egocentric viewpoint has a number of benefits, in particular the ability to refer to the arrangement of features as viewed by the user. This research presents a user centred projective spatial reasoning model which combines and extends previously published projective spatial reasoning models. The proposed extensions improve the fidelity of the model by subdividing projective space into finer addressable units, and through their combination the model is able to summarise relationships between complex objects in 2.5D space, making it suitable for use in spatial projective reasoning queries. The model is demonstrated in a LBS able to establish the visibility of nominated landmarks in a cityscape by using high resolution digital elevation models, which can then support the user who may request information based on the locations of other landmarks (e.g. What's the building on the left of the train station?), or listen to descriptions of landmarks in relation to known features (e.g. the bus stop is in front of the post office). The framework is shown to be able to reason about objects typically in the field of view, and to be suitable for use in spatial queries.

6.2 Introduction

There is no doubt that Location Based Services (LBS) are becoming more commonplace with the increase in georeferenced data (Jiang and Yao 2006) and location aware devices using GPS and network based positioning (LaMarca *et al.* 2005, Skyhook 2008). LBSs are designed for naïve users (Egenhofer and Mark 1995), requiring supportive and 'calming' interfaces (Weiser and Brown 1996) to ensure ease of use in difficult environments. Much of the spatial theory employed by these applications is hidden from the user, such as the ability to use two-dimensional spatial reasoning (in R^2) to determine which features are within defined zones (e.g. finding all hotels within a city boundary). System usability is very important for the success of LBS applications, and great efforts are being made to reduce the seam between the application

and the user by closely modelling the user's viewpoint (Ishii *et al.* 1993, Ishii *et al.* 1994). For example, way-finding applications use the direction of travel to determine the frame of reference by which turning instructions are relevant to the user.

While 2D topology is sufficient to impart navigational instructions, the egocentric viewpoint which LBS users experience is 3D (Meng 2005, Reichenbacher 2005). Communication typically involves turning the relationships between 3D objects into qualitative abstractions (Cohn and Hazarika 2001), often incomplete or inaccurate yet sufficient to convey in linguistic terms the spatial relations between objects (Jackendoff 1992, Jiang and Yao 2006). Language defines space according to an axial structure (Munnich *et al.* 2001), yet in comparison to the number of nouns available only a few spatial terms exist to describe the relationships between objects (Jackendoff 1992). According to Freeman (1975) there are only 13 primitive spatial relationships, as shown in Table 6-1. Primitive relations form the minimum set of descriptors which may be combined to form more complex spatial descriptions, usually limited to a combination of two primitives (Gapp 1995). For example there is no single term for "*left of* and *above*", instead two primitive terms are used in conjunction.

Table 6-1: Egocentric Primitives Spatial Relations - Freeman (1975)

Left of Right of	Above Below	In front of Behind	Inside Outside	Near Far	Beside	Touching	Between

It is argued here that a LBS able to spatially reason from a number of different frames of reference would be able to more closely model the user's environment. This closer integration would lead the way to applications able to filter and describe spatial relationships in a more natural way, such as "take the *first left* and you'll see the supermarket *behind* the park". Such descriptions draw on knowledge of the user's location and direction to determine the appropriate terminology, as well as an ability to determine what is in the user's field of view, and reason what is '*behind'* another feature from that viewpoint. The ultimate goal would be for an LBS to pass the 'spatial Turing test' (Winter and Wu 2008), whereby it's instructions are indistinguishable from those generated by a human. A key element currently missing from published research is a model which unites the output from visibility results within a framework of projective relations enabling the description of object positions in relation to other objects in view (e.g. landmark buildings) as seen from the observer's viewpoint.

The research presented fills this void by combining visual exposure modelling with projective reasoning models to determine which parts of Features of Interest (FOI) are visible to a LBS user, summarising the relationships between FOIs using a projective reasoning model, such that qualitative descriptions are possible. The projective relationship model presented combines a number of existing projective reasoning models (Billen and Clementini 2006), extended through

the adoption of Allen's (1983) interval definitions to increase the range of descriptions. The results are summarised as a tree structure providing a framework to describe the relationships between FOI bodies, and may be serialised making it suitable for storing and querying. The chapter begins with a summary of the existing spatial reasoning models (Section 6.3), before introducing the extensions used for the new combined model in Section 6.4. The model is tested within a virtual city guide LBS which uses visibility modelling to calculate the visual exposure of FOIs as outlined in Section 6.5. The chapter concludes in Section 6.6 with suggestions for how spatial reasoning may be used in conjunction with other techniques to more closely model the user's view for use in LBS.

6.3 Spatial Reasoning Review

There are a number of formal models for spatial reasoning, which in Geographic Information Science are usually based on two-dimensional datasets. The most notable definitions describe the topological relationships between features, such as the 9 intersection model (Egenhofer and Herring 1990, Egenhofer 1991, Egenhofer and Franzosa 1995), Region Connection Calculus (Cohn 1997, Cohn and Hazarika 2001) and the Calculus Based Method (Clementini and Di Felice 1995, Clementini and Di Felice 1997, Clementini *et al.* 1997). The Dimensionally Extended 9 intersection model (9DE-IM) (Clementini and Di Felice 1995) differentiates between the type of intersection (point, line, polygon), and is particularly useful for querying spatial data, leading to its adoption as a standard (OGC 2006). This model is based on topological space, and therefore coordinate independent, resulting in identical descriptions for both scenarios depicted in Figure 6.1.

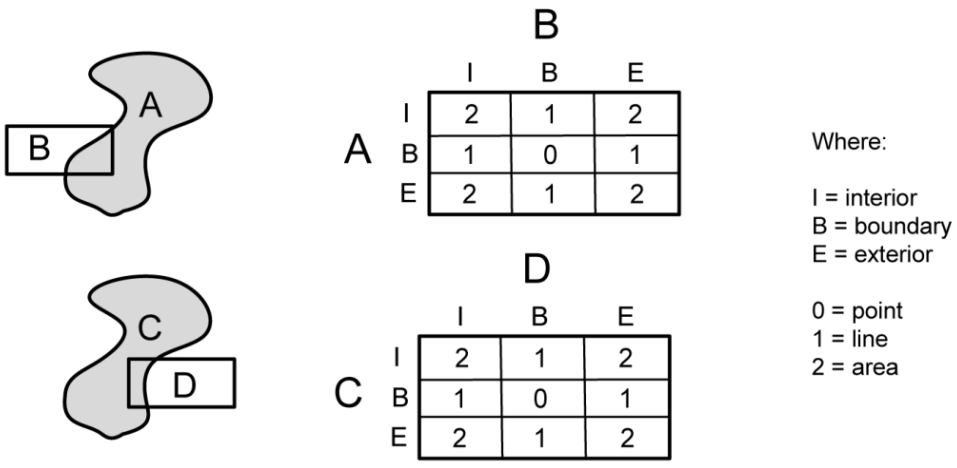


Figure 6.1: Example of 9DE-IM – Region Overlap

Besides topology, egocentric reasoning should consider the user's viewpoint (Hernández 1991, Tarquini *et al.* 2007, Fogliaroni *et al.* 2009); however while topological relations have been researched in three-dimensional space (Borrmann and Rank 2009) only limited research exists on 3D projective relationships (Billen and Clementini 2006). Projective spatial reasoning considers the relations between objects with respect to a frame of reference, and is therefore suitable for use in LBS. Consider the following examples:

Example a) The bus stop is *in front* of the post office.
Example b) Is that the post office *north of* the Church?
Example c) What's the name of the street *left* of the school?

These examples demonstrate the three types of frames of reference by which spatial relationships are usually defined, which are the *intrinsic*, *absolute*, and *relative* frames (Majid *et al.* 2004). The intrinsic frame of reference is defined by a nominated feature, from which the spatial reasoning is carried out. The functionality of a feature, as in the main building entrance, may be used to determine *front* from *back* (Levinson 2003). For example "the bus stop is *in front* of the post office" (Example a) uses the post office to define the frame of reference from which the spatial relationship "*in front"* then has a meaning. Absolute references (Example b) are defined according to an external framework such as a map grid (e.g. North), while relative references (Example c) are able to describe relationships as seen by an observer's point of view.

Following another categorization, frames of reference are divided into two types, *allocentric* and *egocentric*, depending on whether the origin of the frame of reference is placed outside or inside the observer, respectively (Klatzky 1998).

The absolute frame of reference, which ignores the user's location and orientation, makes comprehension more difficult from the observer's viewpoint, and is the least important for LBS applications. The intrinsic view may be significant in some cases, especially when directions are given with respect to specific landmarks. The relative (and egocentric) frame of reference, being a projected view of space from the viewpoint of an observer, is the most relevant to LBS applications. Absolute and intrinsic relationships are usually binary, while egocentric spatial reasoning is always a ternary comparison between the primary object, a reference object, and the observer (Hernández 1991).

6.3.1 Projective Spatial Reasoning

The primitives of Table 6-1 may in part be determined using the 5-intersection model (Clementini and Billen 2006), which is a model for ternary projective relationships. The model is formed on the basis of collinearity of three points, which is the most primitive invariant in projective space. The traditional definition of geometric collinearity only covers points. Its definition is a set of points which fall on a common line, and for a set of three points may be

written as *coll*(p1,p2,p3). The relationship is symmetrical, meaning that the order of points may be changed while maintaining collinearity, so *coll*(p2,p1,p3) would also hold true.

In the 5-intersection model, collinearity has been extended to consider regions such that it may be used to build a qualitative model of space. There are a number of definitions for collinear regions, but for the purposes of egocentric spatial reasoning the collinear_2 definition (Billen and Clementini 2005), as used by the 5-intersection model, is most appropriate. Its definition is that all of the points (*p1*) within the primary feature (A) must be collinear with a point (*p2*) of the observer (O) and a point (*p3*) in the reference object (B). For clarity the definition is rewritten here using O to define the LBS user (i.e. the observer), B as a reference object, and A as the primary object,

$$coll\_2(A,O,B) = \forall p_1 \in A\ [\exists p_2 \in O\ [\exists p_3 \in B\ [coll(p_1, p_2, p_3)]]].$$

This may be explained by considering the 2D case in Figure 6.2a, whereby region A is considered collinear with the view from O to B (referred to from now on as OB). The reference frame is built around O and B using external and internal tangents to define the collinear and aside acceptance zones. Figure 6.2b shows a similar situation except only part of region A is collinear with OB, and is therefore considered partially collinear, and partially aside. The appropriate 5-intersection matrix results for each example are also shown in Figure 6.2.

Collinear space may be refined into between, before, and after through considering the order of regions in the direction of OB. The Aside relations may be refined into Left and Right in R^2 space, thereby creating the 5-intersection matrix (Clementini and Billen 2006).

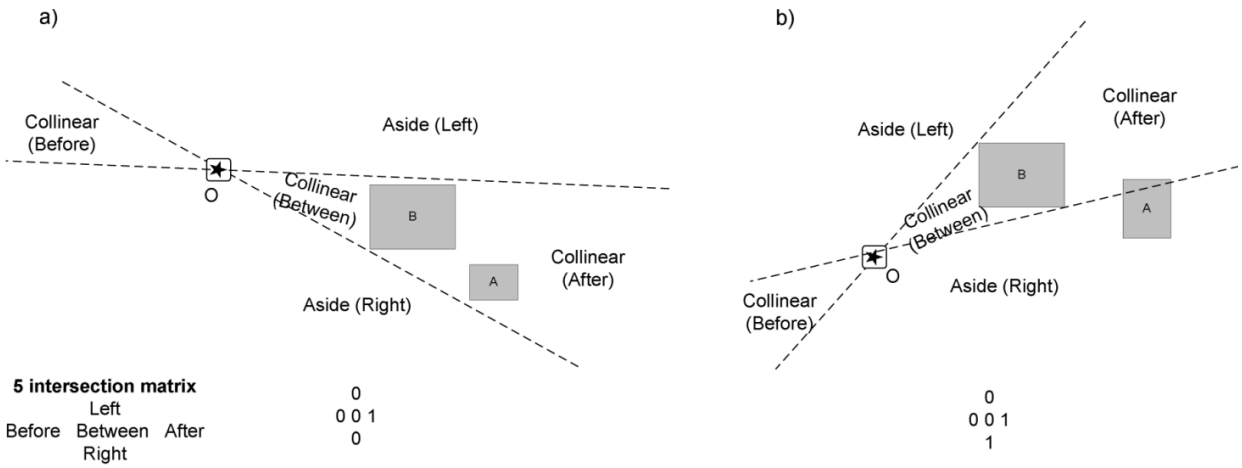


Figure 6.2: Collinear and Aside zones for 1 point and two regions

The model is suited to describe the projected relations for FOI 2D boundaries, but does not consider the 3D aspects of features and therefore additional modelling is required in the form of a quaternary relational model.

6.3.2 Quaternary Relational Model in R^3

The quaternary projective relation model in R^3 (Billen and Clementini 2006), uses three reference objects to define a plane from which the relation of a fourth (primary) object may be determined. The plane divides 3D space into two half-spaces, referred to as HS^{+ve} and HS^{-ve}, which correspond to *above* and *below*. However this does not provide a definition to reason for 3D bodies, and it is therefore necessary to define a coplanarity subspace as presented in Figure 6.3. By defining two planes, one between the observer and the base of the reference object, the second to the top of it, a volume is created which can be used to describe 3D bodies. Anything occupying the space between planes is in Coplanar Subspace (CS) (e.g.1 and part of 2), while anything above the subspace is in CS^{+ve} (e.g. 3 and part of 2); in this example, there is nothing in CS^{-ve}.

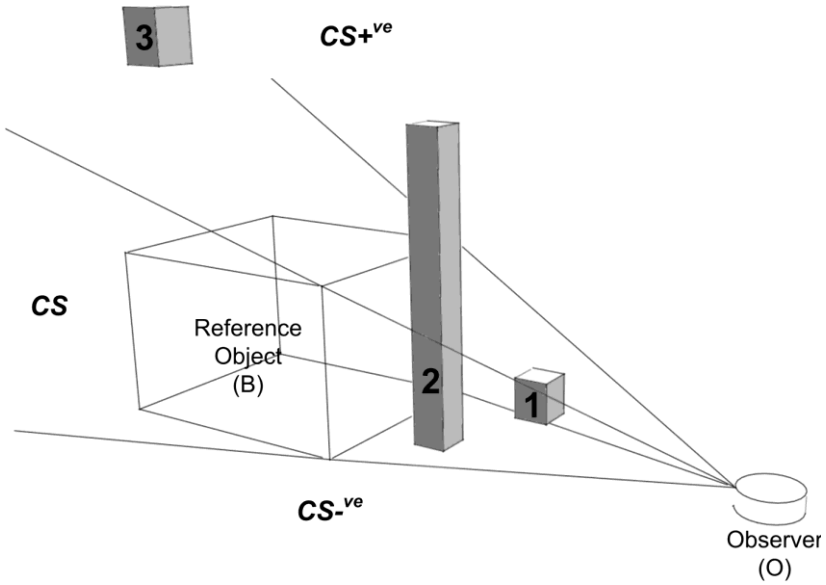


Figure 6.3: Defining Coplanar, Above and Below for 3D Bodies using a Quaternary Model

6.4 A Combined Model, with Extensions

While the quaternary model in R^3 is able to define coplanarity relations among 3D bodies, it is not able to define relationships such as Left and Right. Therefore for the purposes of describing FOI positions to a LBS user it is beneficial to merge it with the 5-intersection model. In addition we propose a number of extensions to increase the fidelity of the output and to use the intrinsic frame of reference, centred on the observer, to describe the space behind the observer which is considered to be out of view. The combined model is intended for use by LBS applications which can access high resolution 2.5D elevation models, and therefore unable to support full 3D volumetric containment concepts. It is however able to describe complex relations between objects, be mapped to Freeman's primitives, and may be used in spatial projective queries.

Figure *6.4* shows how the acceptance zones around the observer are defined according to this approach.

The combined model begins with the premise that when the primary object is behind the observer an intrinsic frame of reference is used, such that the FOI may be described in relation to the observer. By dividing the space behind the observer into 45 degree sectors it is possible to construct sentences such as 'the bank is *behind* you and to your *right*', or 'the bank is on your *right* just *behind* you' giving an indication of the magnitude of direction. In these cases the observer (O) is referenced (*you/your*) which is considered more appropriate than reporting the primary object (A) to be *before* the original reference object (B) as returned from the 5-intersection model, even though it is behind the observer. For any regions which extend across the border from *behind* to *in front* of the observer the term *alongside* may be added to the relation, such that *alongside right* and *alongside left* are used.

In all other cases the primary region is reasoned in relation to the reference object, and its position described using the 5-intersection model. The vertical elements (*above*, *coplanar*, *below*) are reasoned using a quaternary relational model in R^3 for the same reference object, giving a total of 26 (3^3-1) addressable projected zones.

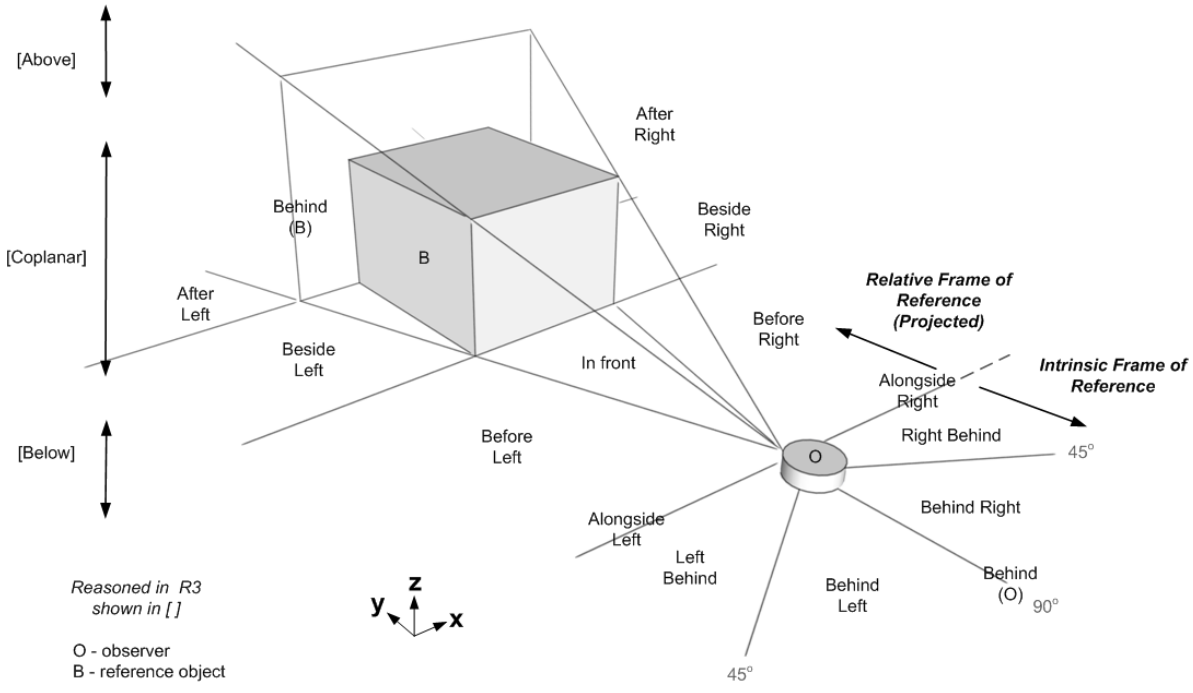


Figure 6.4: Acceptance Zones for the Combined Model

To improve the model's reasoning granularity while conforming to projective geometry invariant restrictions, a number of model extensions are proposed. Currently the 5-intersection model is unable to differentiate between the *Before Left*, *Beside Left*, and *After Left* regions

marked in Figure 6.4, instead describing the result as *Left*. In fact all of the primary regions (A1,A2,A3,A4) in Figure 6.5 are indistinguishable from the results of the 5-intersection model. However if the primary and reference objects are projected onto a line between O and the focal point on B (i.e. where the observer is looking), then a 1D collinear set is available for further reasoning (Figure 6.5). This set is used to order the primary regions from *before* to *after* the reference object, using only relative orders and no numeric measurements. By considering the most extreme points for each region (shown as – and + in Figure **6.5**), it is possible to define a set of distinct outcomes thereby increasing model granularity. The next section describes the implementation of this extension for the combined model considering the case for each axis independently, beginning with an improvement in reasoning fidelity for depth (Y-axis).

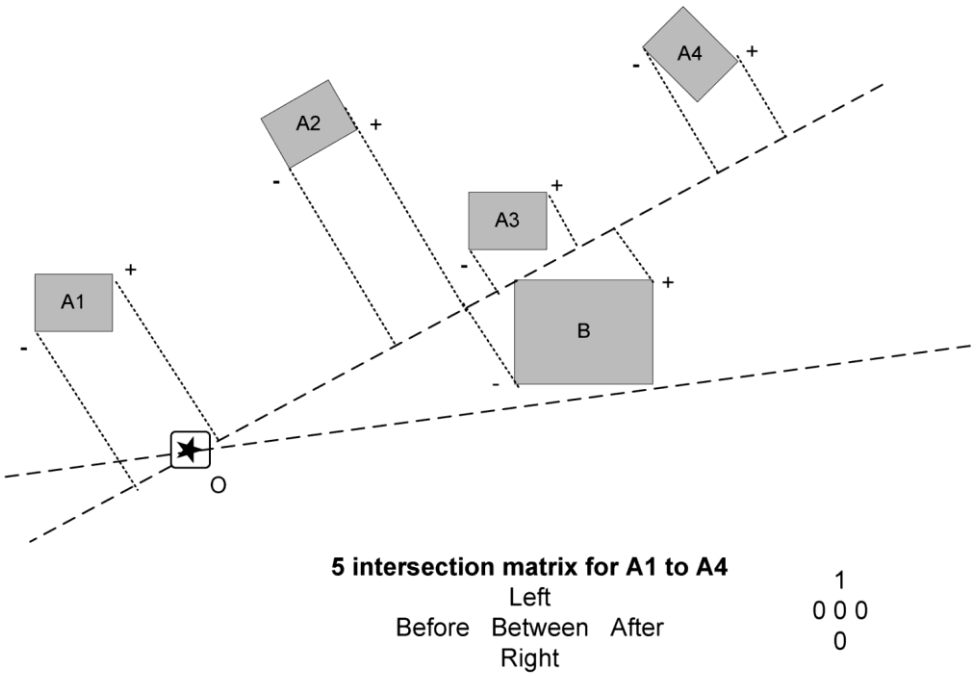


Figure 6.5: Extending the 5-Intersection model for Regions

6.4.1 Graded Cases for Y axis

The 5-intersection model only describes concepts of depth (Y-Axis), using the terms *In front* and *Behind*, for collinear space. Here we propose that *aside* space may also be subdivided through consideration of other models.

Allen's (1983) concepts of temporal relationships may be applied to one dimensional space to describe the relationship between two objects (Güsgen 1989). A region may be defined in 1D space using its extreme points, those that appear first and last on the axis being considered. Therefore by using the extreme points from both the primary region (A) and reference object (B) thirteen identifiable relationship cases may be defined conforming to Allen's 13 temporal

relationships. Figure 6.6 shows the definitions for the Y-Axis which fall into six main classes describing the depth of space from the observer as *before*, *just before*, *nested*, *just after*, *after*, and *alongside.*

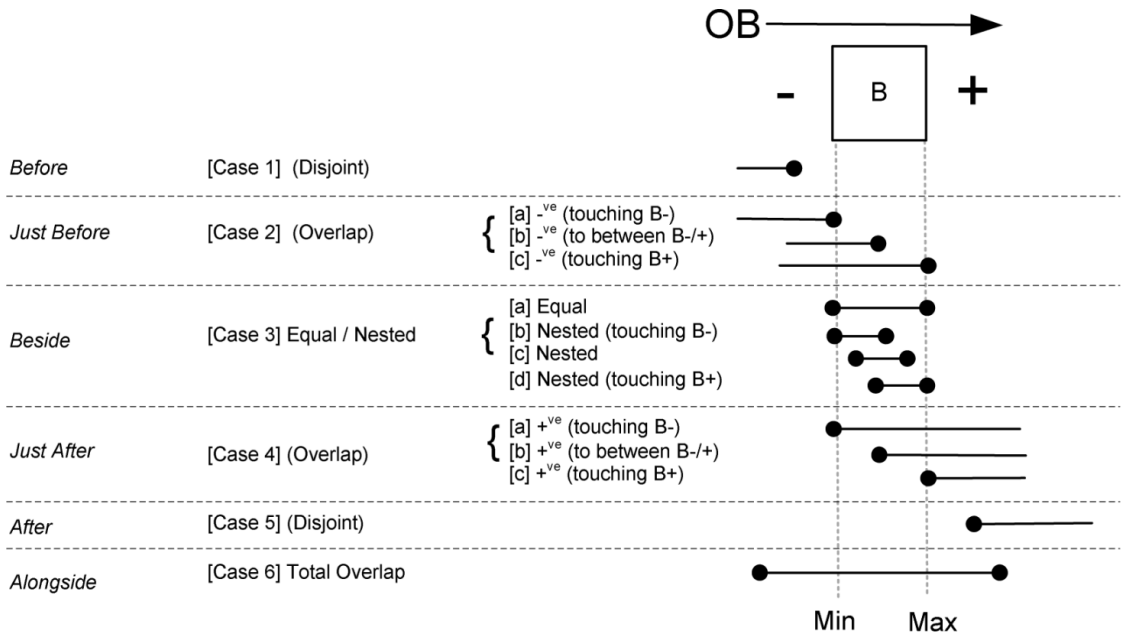


Figure 6.6: Relations in 1D using Overlap and Disjoint refinement (based on Allen 1983)

From this the positions of the primary objects in Figure 6.5 may be described as follows: A1 is *before* and *left* of B, A2 is *just before* and *left*, A3 is *beside* B on the *left*, and A4 is *after* B on the *left*. More specifically these may be referred to as A1 is *Case1*, A2 is *Case2a*, A3 is *Case3c*, and A4 is *Case5* as described in Figure 6.6.

Although collinear space has existing definitions for *in front* and *behind* an additional definition of *nested in front* and *nested behind* space may be introduced, which arises when the convex hulls overlap. For example in Figure 6.7 the primary object is partly left (1) and partly collinear (2,3). The collinear space may be divided into that which is completely behind (2) the reference object (B), and that which is nested (3). To assist with reasoning clarity a differentiation in terminology is made between *before* and *in front*, and *after* and *behind*, for aside and collinear space respectively.

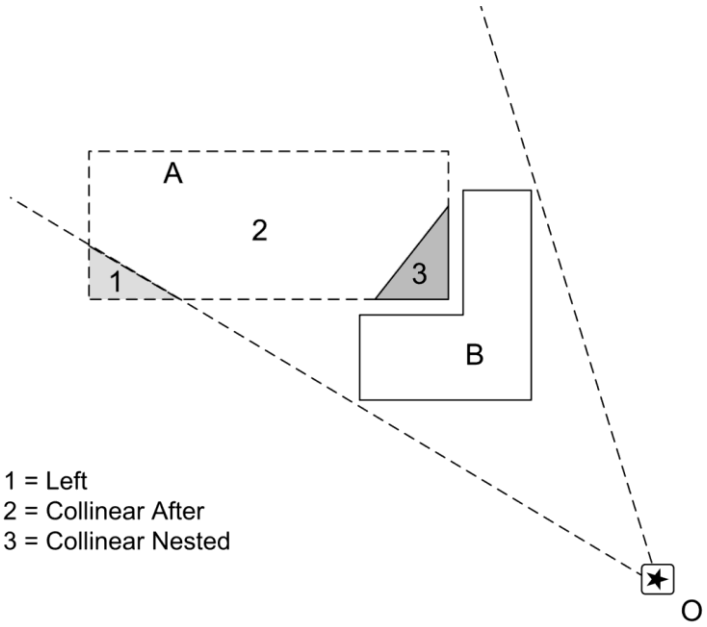


Figure 6.7: Collinear Nested Space

6.4.2 Graded Cases for X axis

A similar refinement may be made to the X-Axis such that objects to the left and right may be defined more specifically. In projective geometry angles may not be used as part of a definition as they are not preserved throughout transformation, however collinearity is invariant. The existing definition for collinear regions (coll2) may be extended to describe the thirteen cases (Figure 6.6) by considering the extreme points from each region. These are the furthest left and furthest right points for a region, denoted using $_{+ve}$ and $_{-ve}$ suffixes. The six main X-axis cases are shown in Figure 6.8a, and a number of graded examples are shown in Figure 6.8b, where A is the primary object and B the reference object. This approach permits the use of terminology such as *immediately right*, *in front just right*, *in front right*, *in front* and *totally in front*.

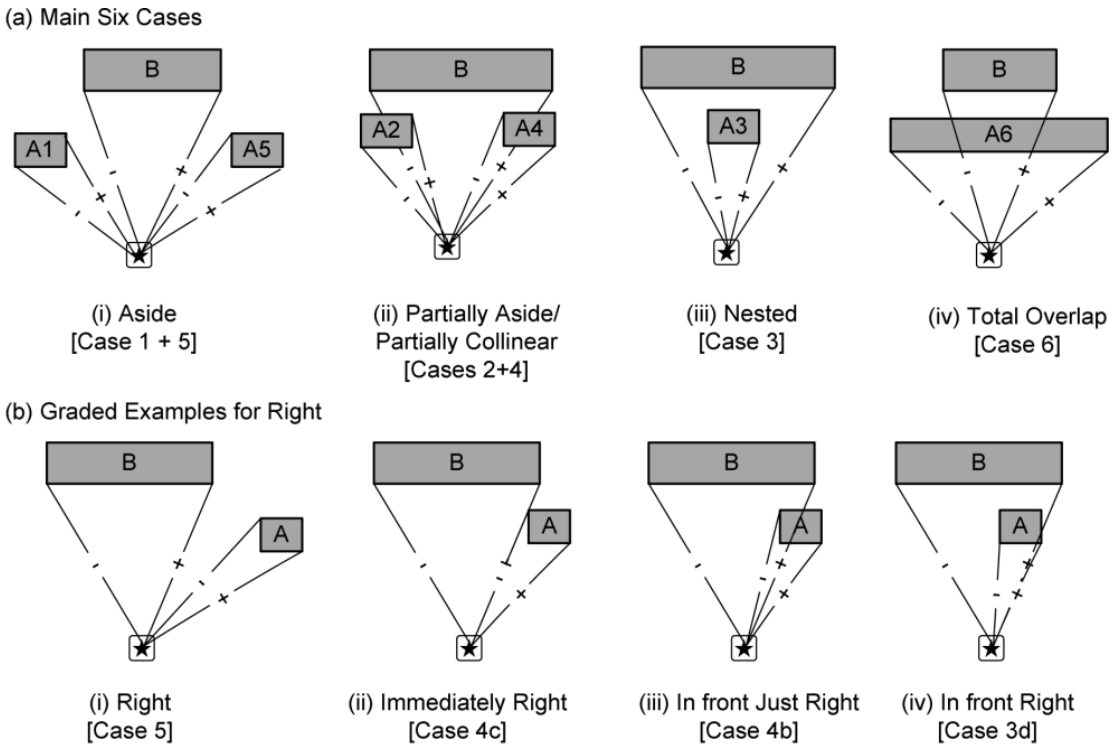


Figure 6.8: Testing for Collinear and Aside Cases where B is the Reference object, A is the primary

The full set of thirteen cases for the X-axis may be defined using tests of collinearity between the extreme points and the regions as shown in Table 6-2. For example Case 1 states that to be true all points (x) from Region A should be left of the zone formed between the observer (point y) and an interior point (z) of convex hull in region B. Case 2b is true when the positive extreme point (x_{+ve}) from region A and the observer (point y) and a point z from the convex hull formed from the interior points of region B should be collinear, and the negative extreme point (x_{-ve}) from A is on the left side of OB. By adopting these extensions a graded approach to the X-axis definitions can be introduced in to the model's definitions.

Table 6-2: Determining X-axis Case based on Collinearity for Visible Regions and Extreme Points

Case1(A,O,B)	$\forall x \in A\ [\exists y \in O\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{ls}(x,y,z)]]]$
Case2a(A,O,B)	$\mathrm{coll}(x_{+ve},y,z_{-ve}) \wedge x_{-ve} \in A\ [\exists y \in (O)\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{ls}(x_{-ve},y,z)]]]$
Case2b(A,O,B)	$x_{+ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{coll}(x_{+ve},y,z)]]]$ $\wedge\ x_{-ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{ls}(x_{-ve},y,z)]]]$
Case2c(A,O,B)	$\mathrm{coll}(x_{+ve},y,z_{+ve}) \wedge x_{-ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{ls}(x_{-ve},y,z)]]]$
Case3a(A,O,B)	$\mathrm{coll}(x_{-ve},y,z_{-ve}) \wedge \mathrm{coll}(x_{+ve},y,z_{+ve})$
Case3b(A,O,B)	$\mathrm{coll}(x_{-ve},y,z_{-ve}) \wedge\ x_{+ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{coll}(x_{-ve},y,z)]]]$
Case3c(A,O,B)	$\forall x \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{coll}(x,y,z)]]]$ $\wedge \neg\mathrm{coll}(x_{-ve},y,z_{-ve}) \wedge \neg\mathrm{coll}(x_{+ve},y,z_{+ve})$
Case3d(A,O,B)	$\mathrm{coll}(x_{+ve},y,z_{+ve}) \wedge x_{-ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{coll}(x_{-ve},y,z)]]]$
Case4a(A,O,B)	$\mathrm{coll}(x_{-ve},y,z_{-ve}) \wedge x_{+ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{rs}(x_{+ve},y,z)]]]$
Case4b(A,O,B)	$x_{-ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{coll}(x_{-ve},y,z)]]]$ $\wedge\ x_{+ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{rs}(x_{+ve},y,z)]]]$
Case4c(A,O,B)	$\mathrm{coll}(x_{-ve},y,z_{+ve}) \wedge x_{+ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{rs}(x_{+ve},y,z)]]]$
Case5(A,O,B)	$\forall x \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{rs}(x,y,z)]]]$
Case6(A,O,B)	$x_{-ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{ls}(x_{-ve},y,z)]]]$ $\wedge\ x_{+ve} \in A\ [\exists y \in \mathrm{O}\ [\exists z \in \mathrm{CH}(B^\circ)\ [\mathrm{rs}(x_{+ve},y,z)]]]$
Coll = collinear LS = left side RS= right side CH = convex hull *x,y,z are points in regions AOB respectively* *Extreme points are denoted by suffix $_{-ve}$ or $_{+ve}$*	

6.4.3 Graded Cases for Z axis

Previously, the concept of quaternary relationships for determining *above* and *below* in R^3 was raised in Figure 6.3. An example for a more complex primary body is given in Figure 6.9, whereby a number of planes are created between the observer and reference object. The primary object (A) being considered as a set of individual columns for which the appropriate Z-axis case may be determined, with the overall classification accounting for the range of cases encountered. In Figure 6.9 the primary feature (A) would be classed as *Case 4a*, signifying it occupies the coplanar subspace and extends into CS^{+ve}.

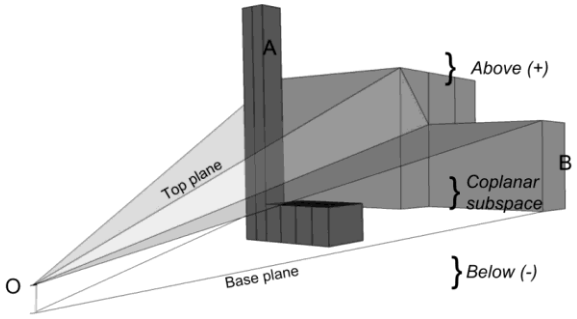


Figure 6.9: Spatial Reasoning for Z-Axis

The methods described in Sections 6.4.1,6.4.2,6.4.3 are used to reason between a primary and reference object in 2.5D, with results being stored in a tree form as described in the next section.

6.4.4 Reasoning Summary – Tree form

The results from reasoning for each axis may be stored according to the thirteen cases shown in Figure 6.6 using a ternary system, whereby 'F' indicates an empty set, '0' indicates extreme points of primary and reference objects are collocated on an axis, and '1' indicates a line passes through the region as shown in *Figure 6.10*.

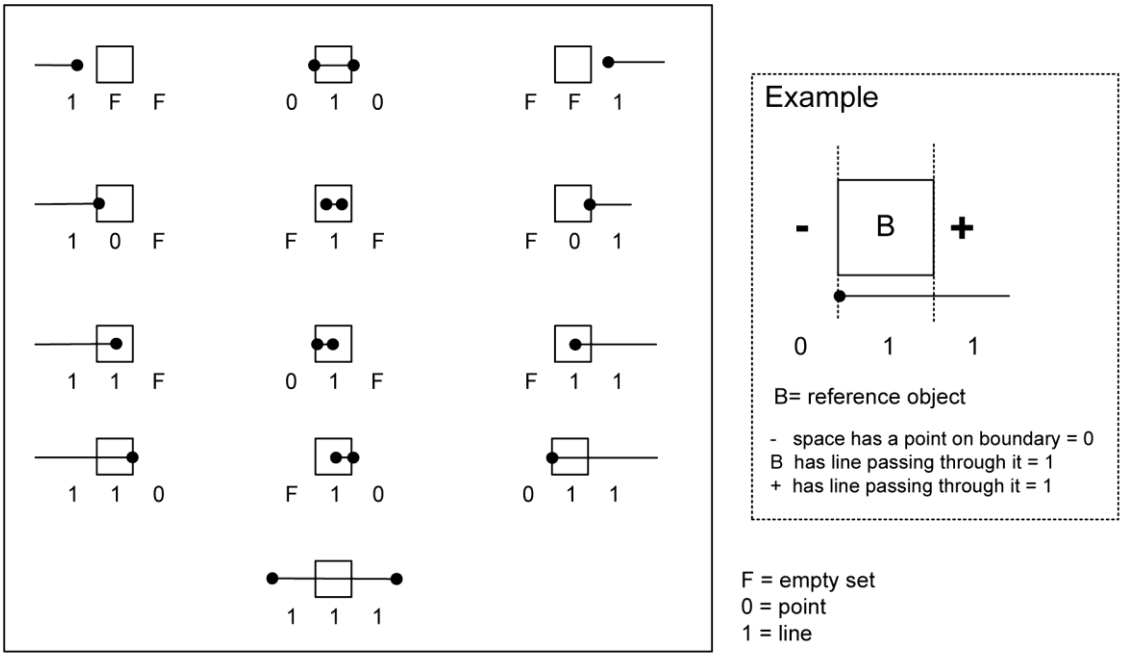


Figure 6.10: Reasoning Refinements

The order of reasoning between objects is as follows: reason for X axis to determine left, collinear, and right sections of the primary object. Each section is then reasoned again for the Y axis creating sub-parts which are reasoned in the Z axis, forming a tree as displayed in Figure 6.11, which may be serialised. Here the tree is serialised by depth with a decimal point placed between each branch and a colon between each layer to improve reading ease. Square brackets added to subparts of the tree assist when searching for specific patterns in the abbreviated serialised form, as becomes evident later in the chapter. Values from lower branches are only included if a True condition is met, removing excessive False nodes from the serialised result. This single string is able to describe all possible projective relationships a solid primary object may have with a solid reference object in 2.5D space. A number of examples for simple and complex shapes are shown in Figure 6.12.

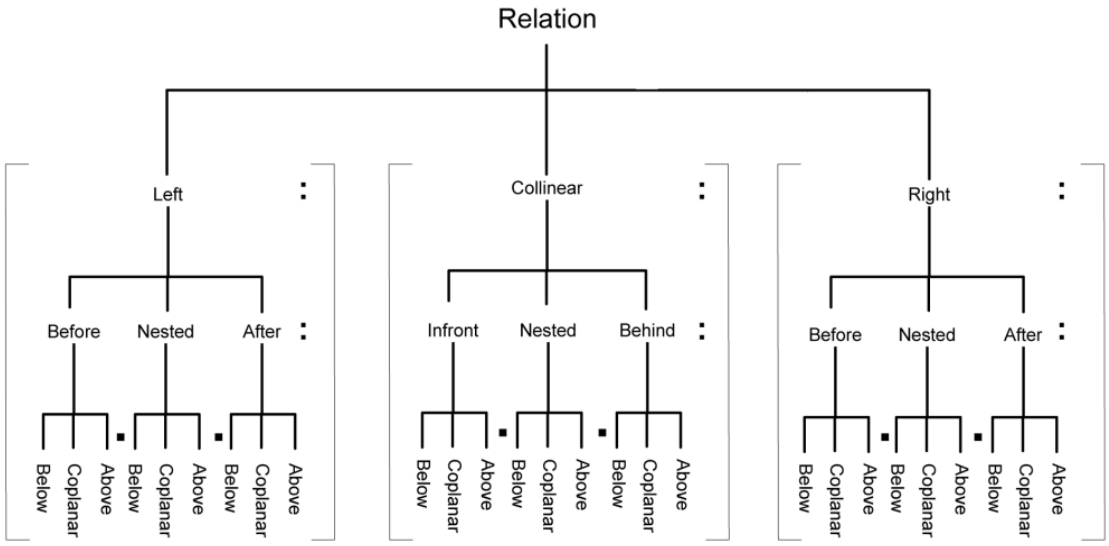


Figure 6.11: Projective Reasoning Tree

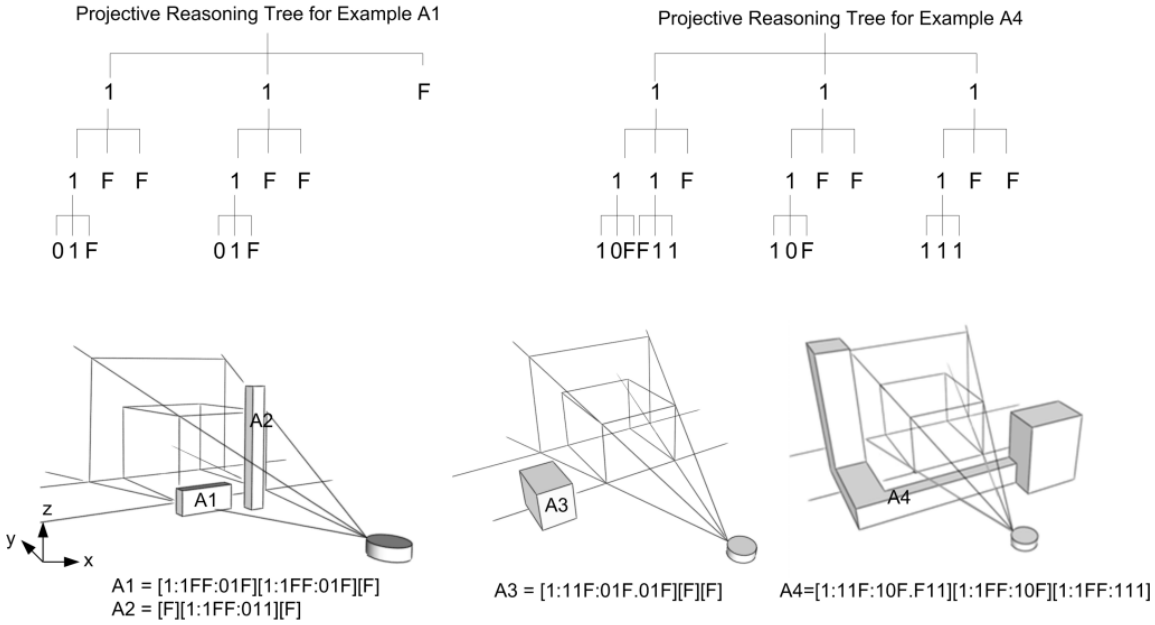


Figure 6.12: Reasoning Tree Examples

The serialised value may be used to describe projective relationships between the visible aspects of two objects in 2.5D space, without the concept of volumetric containment, suited to the egocentric reasoning required for LBS. Where specific criteria are not required a wildcard (*) may be used denoting any sequence may be substituted into that portion of the serialised key, a question mark (?) where any single character (F,0,1) may be used, or the letter T when a True value (0 or 1) is required. For example the search term [1:F1F:*][F][F] would match any primary object which is *beside left* of the reference object. All items *Above* the reference object

could be found using the search criteria [*:*:F?1], while those *Below* would be found using [*:*:1?F].

6.4.5 Implementation of Freeman's Primitives for 2.5D LBS

The model presented combines existing theory with a number of extensions for the purpose of projective reasoning for LBS in 2.5D, and is based on projective invariants without the use of metric distances or angles. The model is able to describe complex relationships beyond those used in language, and therefore it is useful to map the outputs to the Freeman's Primitives. As shown in Table 6-3 this is possible for the majority of cases with the exception of the definitions of *Near* and *Far* which require a definition using metric space. Other terms such as *Inside* and *Outside*, and *Touching* may reference the DE-9IM, as topological relationships hold in projective space.

Near and *Far* are fuzzy distance descriptions and therefore more suited to metric based reasoning. There is no exact boundary, and a reference scale is critical in determining the relationship (Peuquet 2002). For example two buildings may be described as near each other while two cups on a table may not be. To overcome this scale issue Abella and Kender (1993) define the *Near* relationship to be when the bounding boxes of objects have non-empty intersections. *Far* is defined as when the distance between bounding boxes is greater than the larger of the bounding box's longest axis. According to this definition two objects may be neither *near* nor *far* from each other.

The term *Beside* may be defined as the zone of nested space in the Y-axis and aside space in the X-axis, as shown in Figure *6.4*. However the term implies a finite zone near the reference object and not an open set as currently defined. To limit the region's extent the *Near* relationship may be included in the definition, using metric space as outlined above. However an alternative approach using projective relations would be to use the reciprocal relations of AOB and BOA as follows. If a part of region A falls within the *Beside* relation of B, and part of region B falls within the *Beside* zone of A, then the two regions may be considered to be *Beside* each other. For example in Figure 6.13 the *Beside* zone for the reference object is denoted by the B- and B+ lines, which includes both primary objects A1 and A2. The *Beside* zone for A1 includes B, and therefore these two objects are considered to be *Beside* each other, whereas A2 is not considered to be Beside B. This method accommodates the scale of the objects such that the definition is suitable for small or large objects (e.g. buildings or mountains).

If the user approaches and enters the zone between the primary and reference buildings the projective relationship would change to a Left of / Right of, or In Front / Behind (you) relationship depending on the observer's viewing direction. The relationship with regard to the intrinsic frame of reference would remain unchanged.

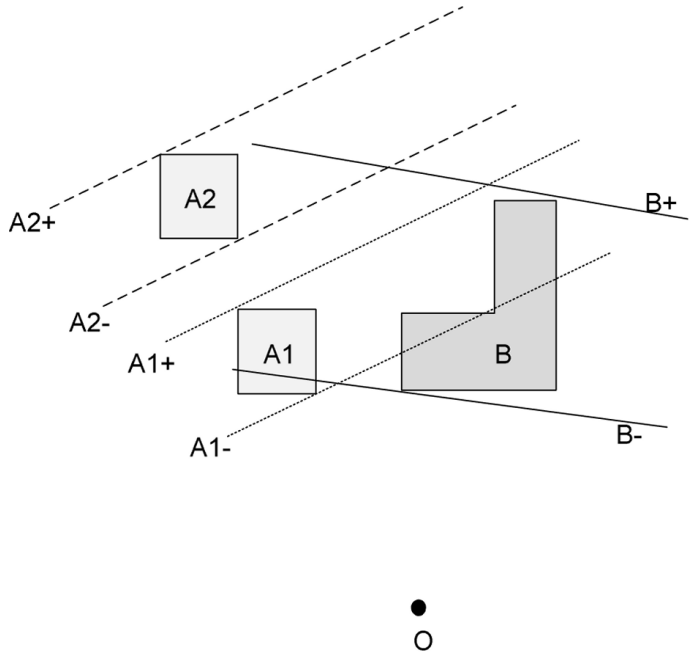


Figure 6.13: Beside Relation defined using a Reciprocal Reasoning test

Table 6-3 shows examples from the serialised tree, where a wildcard (*) implies any result combination may be substituted for that part of the reasoning tree, and T indicates a true value (0 or 1) is required.

The next section demonstrates the model's use in a real world scenario, whereby the reasoning model is able to define and find FOIs which are visible to a LBS user.

Table 6-3: Implementation of Freeman's Primitives

Primitive Relation	Graded Extensions	Search Criteria Examples
Left of	Left	[1:*:*][F][F]
	Immediately Left	[1:*:*][0:*:*][F]
	Just Left	[1:*:*][1:*:*][F]
Right of	Right	[F][F] [1:*:*]
	Immediately Right	[F][0:*:*][1:*:*]
	Just Right	[F][1:*:*][1:*:*]
In front of	In Front of	[F][1:1FF:*][F] [F][1:10F:*][F]
	In Front Left	[1:1FF:*][0:1FF:*][F] [1:10F:*][0:10F:*][F]
	In Front Right	[F][0:1FF:*][1:1FF:*] [F][0:10F:*][1:10F:*]
	In Front Just Left	[1:1FF:*][1:1FF][F] [1:10F:*][1:10F][F]
	In Front Just Right	[F][1:1FF:*][1:1FF:*] [F][1:10F:*][1:10F:*]
Behind B (reference object)	Behind	[F][1:FF1:*][F] [F][1:F01:*][F]
	Behind Left	[1:FF1:*][0:FF1:*][F] [1:F01:*][0:F01:*][F]
	Behind Right	[F][0:FF1:*][1:FF1:*] [F][0:F01:*][1.F01:*]
	Behind Just Left	[1:FF1:*][1:FF1:*][F] [1:F01:*][1:F01:*][F]
	Behind Just Right	[F][1:FF1:*][1:FF1:*] [F][1:F01:*][1:F01:*]
Behind O (observer)	Relative position calculated in metric space and graded into 7 zones	
Above	Above	[*:*:FF1]
	Immediately Above	[*:*:F01]
	Just Above	[*:*:F11]
Below	Below	[*:*:1FF]
	Immediately Below	[*:*:10F]
	Just Below	[*:*:11F]
Beside	Beside Left (holds for AOB &BOA)	[1:F1F:*][F][F]
	Beside Right (holds for AOB &BOA)	[F][F] [1:F1F:*]
Touching	Use DE-9IM as topological relationships hold in projective space	
Between	Quaternary projective relations between bodies in R^2 Between(AOBC)	
Near / Far	Metric space model [37]	
Inside / Outside	Use DE-9IM topological model (e.g. the stream is inside the park)	

6.5 Application of Combined Model Projective Reasoning

6.5.1 Visibility Analysis

The model's implementation requires an ability to calculate what is in the user's field of view. This may be done using visibility modelling which in turn requires a Light Detecting and Ranging (LiDAR) sourced digital surface model (DSM) to provide data on the city's profile, including buildings, vegetation, and topography (Palmer and Shan 2002, Rottensteiner and Briese 2002, Omasa *et al.* 2008). The visual exposure of each FOI is calculated by considering the lines of sight between the observer and each raster cell within the FOI boundary, considered as a target. The model returns results for each of these targets denoting the extents visible as vertical columns (Figure 6.14) so that various visual exposure summaries may be calculated per FOI, including the total façade area visible. These results may be used to determine, for example, if the entire façade of a church is visible or just the top part of the spire. As the observer moves so the model is re-run giving a dynamic account of the parts of FOIs visible, which can then be summarised using the projective model outlined in this chapter. A more detailed account of the visibility model implemented may be found in Bartie *et al* (2008).

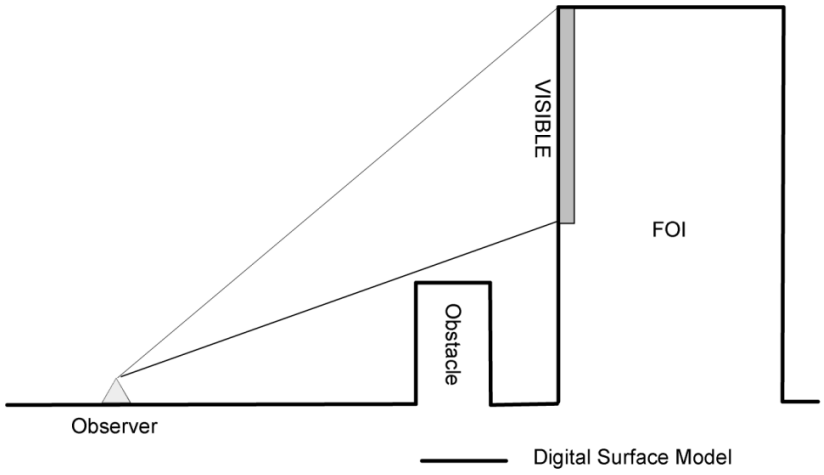


Figure 6.14: Visual Exposure Modelling for a Feature of Interest (FOI)

6.5.2 Case Study 1

A dataset for the city of Christchurch, New Zealand was used to trial the combined projective reasoning model. A DSM at 1 metre resolution was used to carry out the visual exposure modelling for selected FOIs, the results of which quantify the vertical visible extent of the cells within the defined FOI boundary.

Figure 6.15 shows the result from the trial, with the model's description of the relationship between the primary building (A) and the reference object (B). The rays cast by the visibility model are shown from a perspective view in Figure 6.15a, with blue lines indicating the rays from the observer to FOI A, and green lines show the rays to FOI B. A plan view of the situation

is shown in Figure 6.15b, and a photograph with the denoted sections in Figure 6.15c. The results describe which parts of the FOIs are visible to the observer, and define the bodies which are used in the reasoning process. The model tree is completed by first considering the X-axis definitions (i.e. A is collinear, and right of B), then each section is further refined to populate the reasoning tree (Figure 6.11).

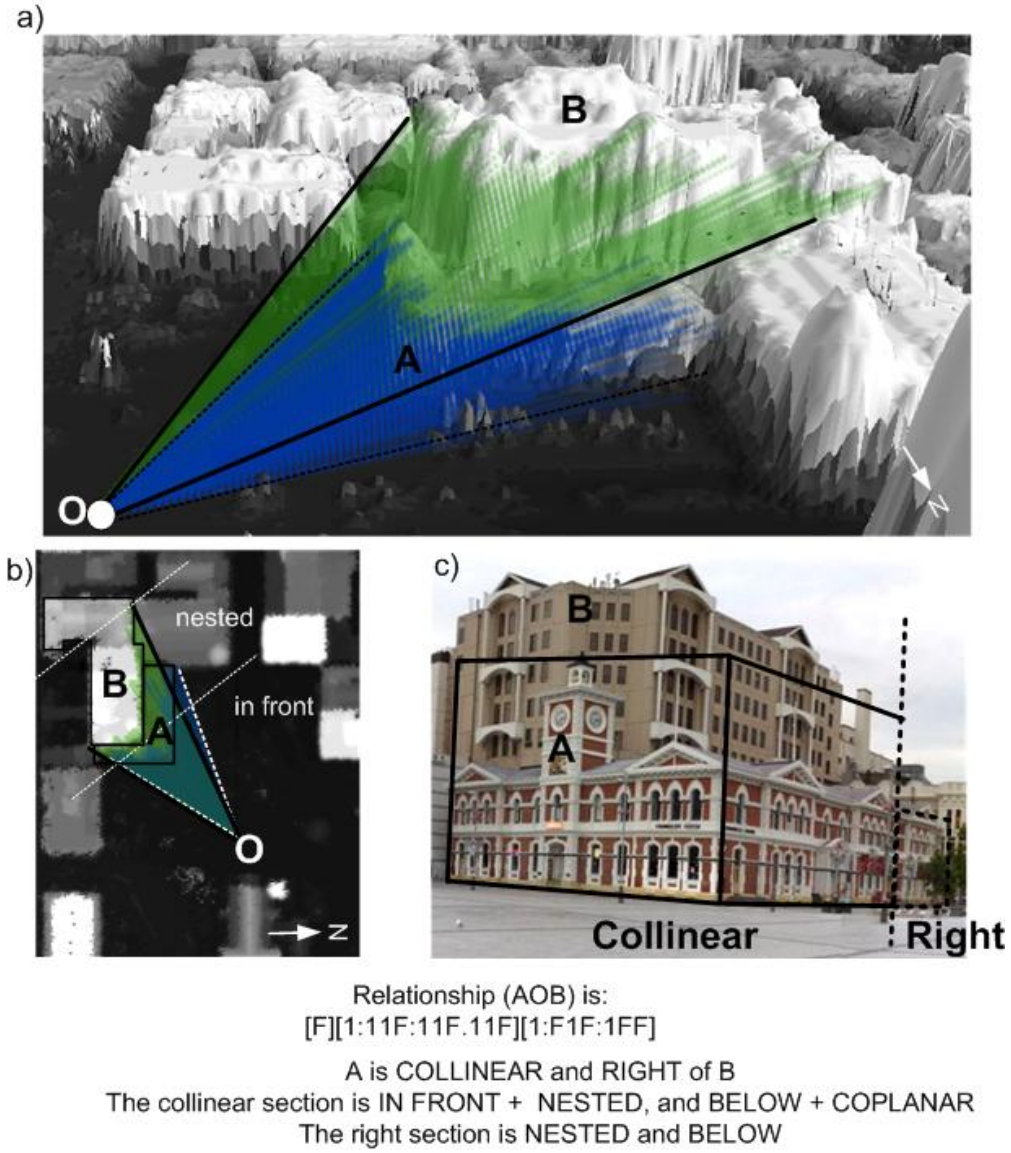


Figure 6.15: Real World Trial – Case Study 1 (a) Perspective View of Visibility Model on LiDAR sourced DSM showing LoSs (b) Plan view of LoSs (c) Photograph from viewing location

A further set of trials were then undertaken to study a more complete set of relations between FOIs.

6.5.3 Case Study 2

For this trial the relationships between a number of FOIs were studied from a single location in Christchurch (NZ), the corresponding photograph and mapped visibility results from which are

included in Figure 6.16. A complete set of reasoning was performed between FOIs C to J, from an observation point in Cathedral Square, with each FOI being used as the reference object. For reasons of space the results in Table 6-4 are abbreviated, showing a complete set of results when C was the reference object, and a summary of the more interesting results from other trials.

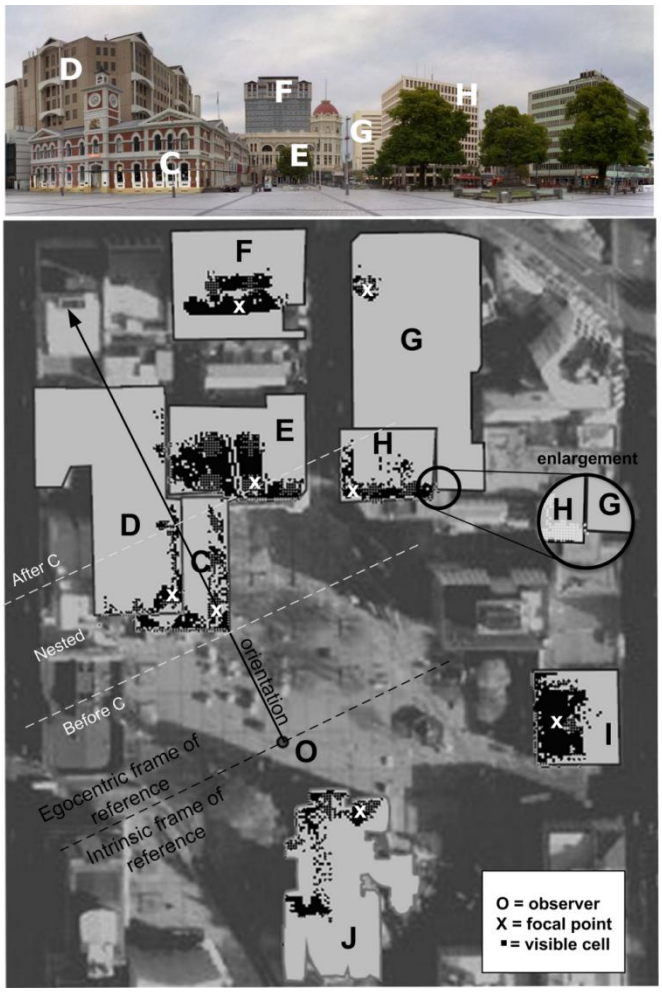


Figure 6.16: Projective Relations – Map showing visibility of buildings from point O. Notations for the example where C is the reference object are also added.

Table 6-4: Egocentric Reasoning Results when looking towards focal point on each Reference Object

Reference Object (B)	Primary Object (A)	Serialised Relation	Explanation
C	D	[1:F1F:F11][1:F11:F11.F11][F]	x: left + collinear y: nested,nested+behind z: coplanar+above, coplanar+above
	E	[F][1:FF1.FF1][1:F11:F1F]	x: collinear + right y: behind,nested+after z: above, coplanar
	F	[F][F][1:FF1:F11]	x: right y: after z: coplanar + above
	G	[F][F][1:F11:F1F.F1F]	x: right y: nested + after z: coplanar
	H	[F][F][1:F11:F11.F11]	x: right y: nested + after z: coplanar + above
	I	Intrinsic	Intrinsic: Right + Behind
	J	Intrinsic	Intrinsic: Behind
E	C	[1:1FF:F11][1:1FF.F11][F]	x: left + collinear y: before,in front z: coplanar+above
F	G	[F][F][1:11F:11F.11F]	x: right y: before + nested z: below + coplanar
G	F	[1:F1F:F11][F][F]	x: left y: nested z: coplanar+above
H	C	[1:1FF:F1F][F][F]	x: left y: before z: coplanar
I	J	[F][F][1:1FF:F11]	x: right y: before z: coplanar + above
J	I	[1:11F:F1F.F1F] [F][F]	x: left y: before + nested z: coplanar

6.5.3.1 Analysis of Results

For the majority of cases the reasoning is fairly simple, however a number of results require further investigation. The relationship between regions is not symmetrical, as demonstrated by *r*(COE) and *r*(EOC). The X-axis relationship is mirrored, with E appearing *right of and collinear* with C, and C *appearing left of and collinear* with E. However E is *nested and after* C, while C is simply *in front* of E. Part of the explanation can be seen in Figure 6.5, whereby A3 is *nested* with B, but B would be considered *behind, nested and in front of* A3. However the other factor is that the depth perception is calculated by using a single focal point, and as this moves between reference objects so the descriptions of depth change.

A further phenomenon of reciprocal projective relationships may be noted in the Y-axis where *r*(IOJ) and *r*(JOI) both report the other to be in the *before* zone. This can be explained with a simple example of two objects on a table, as shown in Figure 6.17. The metric distance OA is greater than OB, as shown by the arc C of radius OB, however from a projective point of

view the relationship is that A is *aside and before* B which agrees with the egocentric view shown in Figure 6.17a. When A is considered as the reference object then B is considered *aside and before* A, demonstrating one of the differences when reasoning in projective space and metric space.

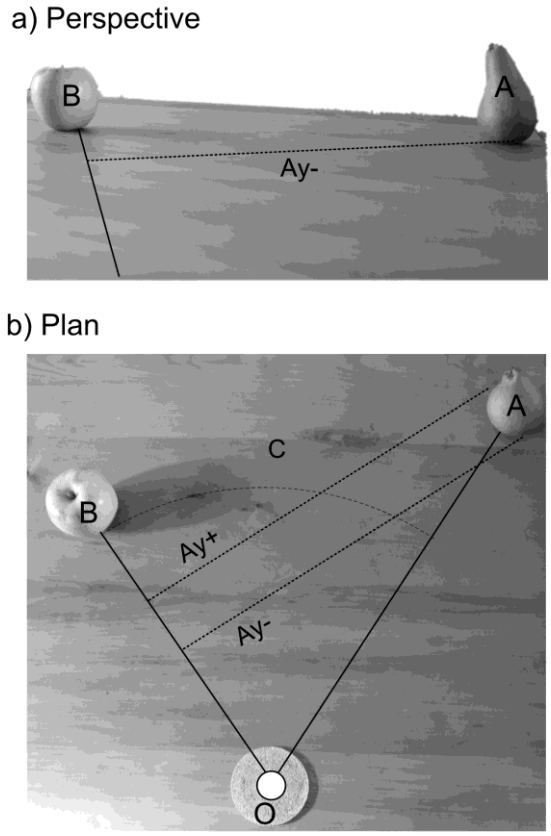


Figure 6.17: Y-axis reasoning - how two objects can share a reciprocal in front relationship

The FOI pairs C-D, D-E, D-H, E-H and F-G are considered to be *Beside* each other according to the reciprocal definition outlined in Section 6.4.5, as demonstrated in Table 6-4 with results F-G and G-F sharing a mutual Y-axis nested relationship. The pairing of D and H is perhaps questionable, but the large size and depth of building D accounts for why the model produces this result.

The result from such projective reasoning may be used to form sentences which describe to a user the position of a FOI in relation to one already known, particularly useful for speech based interfaces. It is also possible to define a spatial query which returns all the regions which satisfy a particular projective relationship, as demonstrated in the last case study.

6.5.4 Case Study 3 – Spatial Queries

The model may be used within spatial queries to determine which FOIs take part in a specified projective relation. A question mark (?) may be substituted for any single value (F,0,1), and an asterisk (*) for any result combination. For example referring to Figure 6.16, a search for the

FOIs *behind* ([*][1:F?1:*][*]) the reference object E, gives a single result of F, while the more general search for those *after* ([?:F?1:*]) gives the results D,E,F and G. A similar distinction may be made between the FOIs which are *right* of E ([*][*][1:*:*]) yielding results G and H, and those on the *right* which extend *after* E ([*][*][1:F?1:*]), returning only G. The FOIs on the *left* may be divided into those *coplanar and extending above* ([1:*:?T1][*][F]), and those totally *above* ([1:*:FF1][*][F]) giving results C and D respectively.

The model may also be used to find the locations where a specific projective relation between two FOIs exists, such as finding where E appears to completely overlap (X-axis *Case 6*) the reference object F. This is done using the search criteria [1:*:*][1:*:*][1:*:*] and gives the result mapped in white in Figure 6.18.

As FOI E has a tower on the right side (as seen in Figure 6.16) a further study was carried out to determine the locations from where this tower would appear to extend *above* FOI F. The query used was [*][*][1:*:??1] which keeps only those locations that have a *right* side component above F. The matching locations are shown in Figure 6.18 as black triangles. This kind of analysis, involving the relationships between FOIs, lends itself to tourist guide LBS applications which would be able to instruct a user to move to certain locations to obtain particular views of the city. It may also have uses in town planning and urban design.

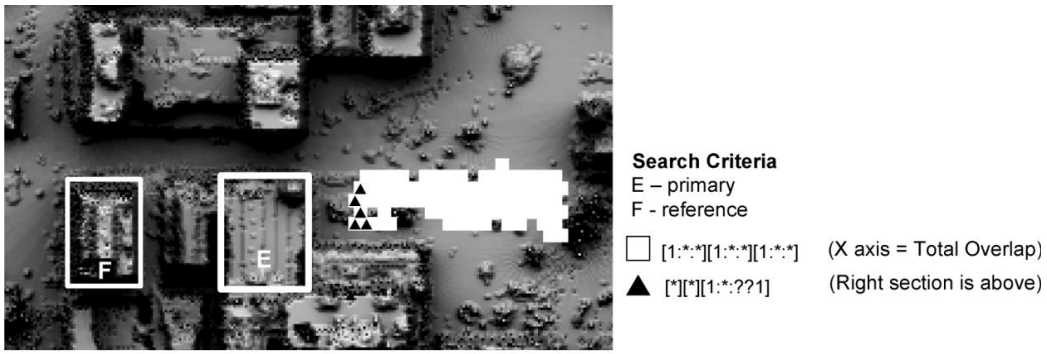


Figure 6.18: Locations which match a given Projective Spatial Reasoning Query

6.6 Conclusion and Future Work

Cognitive models of space are qualitative (Hernández 1991), and applications which are able to adopt human-like reasoning methods are desirable as they imposed a smaller cognitive load on the user. The focus of this research has been to establish a combined projective reasoning model for use in LBS which may be serialized into a single descriptor able to convey complex relationships between reference and primary objects in 2.5D from a specified observation point. The combined model uses a number of existing projective models which were extended by adapting Allen's (1983) temporal model to improve the fidelity of relation definitions. This was achieved using collinearity tests, the only invariant of projective space, allowing the use of phrases such as *immediately right*, *just right*, and *right*.

A number of case studies were used to demonstrate how a high resolution LiDAR based DSM could be used in conjunction with a visual exposure model to establish which parts of FOIs were in view, from which the projective reasoning model was able to describe the relationship between the FOIs as seen from the observer's frame of reference. Such reasoning would be useful in dialogue based systems, allowing the user to construct questions about unknown features by describing their relationship to known ones.

The model may be used to describe the relationship between objects, to search for buildings which match a given relationship criteria, or to find locations where a particular relationship exists between objects. The model may be used in combination with existing spatial and attribute searches, such as to find the names of all buildings of historical significance in view to the right and behind the train station.

6.6.1 Future Work

Further work should consider the issue of primary object fragmentation, which is a concern when a more distant low primary building extends either side of the taller closer reference object. In this case the model describes the relationship as an overlap situation (*Case 6*), however the corresponding details from the collinear section are missing as they are out of sight.

There would be benefits in considering how projective reasoning trials may be combined such that a more comprehensive description of space is formed. For example describing C and D as right of B, then reasoning between C and D to determine which is closer. There may also be benefit in including fuzzy classification methods which consider the proportions of buildings within each zone so that the most dominant classification is used first when describing the relationship. Furthermore projective spatial reasoning may be combined with other datasets, such as topography and building geometry, to generate more complete descriptions, such as "the bank is the *tall* building on the *right* of the *hill*".

Chapter 7: Referring Expressions in Location Based Services: The case of the 'Opposite' Relation

7.1 Summary

Mobile devices and location based services enable digital and real worlds to be integrated within our daily lives. They largely rely on graphical interfaces which require the user to re-focus their attention on the device in favour of their environment, however speech interfaces offer an alternative interaction experience. Aside from speech recognition and synthesis, the handling of natural language dialogue raises several research challenges, including the ability to direct the user's attention to a particular feature in the field of view through the use of suitable descriptions. To mimic natural language these referring expressions should use attributes which include factors of the building's appearance, and descriptions of its location with reference to the observer or other known buildings in view. This research focuses on one particular positional case used in describing features in a field of view, that of the "opposite" spatial relation, and discusses how this referring expression may be generated by modelling the view from the observer's location to the surrounding features. Consideration is also given to the relationships with surrounding entities to find the most suitable reference candidates, and minimise the risk of confusion in cluttered environments. Two models are presented which may be used to determine opposite entities with respect to a linear feature and a regional feature. The research shows that linear feature cases may be more easily modelled than regional cases.

7.2 Introduction

Increasingly digital and real worlds are becoming integrated within our daily lives, with mobile devices and location based services being among the tools that enable this to happen. One of the drawbacks has been that graphical interfaces distract the user from their environment, and alternative interaction experiences are being researched. Augmented Reality (Feiner 1993, Azuma 1995, Hollerer *et al.* 1999, Narzt *et al.* 2006) is one such innovation whereby digital information is superimposed onto real world views. Speech interfaces are another solution, whereby information may be retrieved using voice commands and speech prompts (Holland *et al.* 2002, Goose *et al.* 2003, Du and Crestani 2004, Bartie and Mackaness 2006). The work presented here discusses how speech interfaces may reference items in the current view using natural language terms, with particular focus on the use of the spatial preposition "opposite" as in "we're the house *opposite* the bakery".

People use language to describe and share experiences about space and the objects which occupy it (Jackendoff 1992). These object descriptions are known in natural language research as "referring expressions" and are used, for example, to draw someone's attention to a particular

building in a cityscape (Dale and Reiter 1995). Typically the descriptions include a number of physical attributes relating to the feature, such as its position relative to the observer or other surrounding objects, so that the listener may identify the intended target. The research area has particular relevance for the future of speech based interfaces for Location Based Services (LBS), in both the generation of phrases to direct the user's attention, and in the parsing of phrases to determine which object is being interrogated.

This chapter discusses the positional case of "opposite", and how a LBS may make use of the term when forming referring expressions. To be able to define the case whereby two buildings are considered opposite, the visibility from the observer to each building needs to be determined, as does the view of a common reference point between the two target buildings, for example a road. A model is proposed which makes use of an urban Digital Surface Model (DSM) to estimate visibility, along with a number of examples of how the term may be calculated and used when generating spatial descriptions.

7.3 Background

The range of LBS applications has diversified from basic navigational support into areas of social networking and virtual city guides (Long *et al.* 1996, Abowd *et al.* 1997, Espinoza *et al.* 2001, Bartie and Mackaness 2006). User interfaces have tended to have a graphical focus, but mobile speech recognition tools are improving to the point where devices in the near future may incorporate a speech based mode allowing users to operate in a hands-free and eyes-free way discretely, without the need to re-focus attention from their environment (Francioni *et al.* 2002). Earlier research has shown that users are able to carry out multiple tasks more easily if they are using a number of sensory modalities (Allport *et al.* 1972)

System usability is very important for the success of LBS applications, and great efforts are made to reduce the seam between the application and the user by closely modelling the user's viewpoint (Ishii *et al.* 1993, Ishii *et al.* 1994).The ultimate goal would be for an LBS to pass the 'spatial Turing test' (Winter and Wu 2008),whereby its instructions are indistinguishable from those generated by a human. Steps towards this goal require that the LBS filter and translate digital information into appropriate forms which match the user's frame of reference (Meng 2005, Reichenbacher 2005).

7.3.1 Positional information

The position of any feature in the urban landscape can be described relative to the observer, or a secondary reference object. Relationships are rarely described in metric space (e.g. 123.7m at 54 degrees) but instead usually refer to topological space (Egenhofer and Herring 1990, Egenhofer 1991, Clementini and Di Felice 1995, Egenhofer and Franzosa 1995) or projective space (Clementini and Billen 2006). For example a paddling pool may be described as being "inside the park" using a topological relationship, or "in front of the swings" using a projective relationship. Equally a house may be described as "on your left" by referencing the view experienced by an observer at a given location and orientation.

The topological relationships between static features are permanent, which in urban areas may include containment within a region (e.g. in a park), topographic feature (e.g. on a hill, a slope, or in a valley), or adjacency to a linear feature (e.g. road, river, rail). In contrast projective relationships are ternary comparisons between the primary object, a reference object, and the observer (Hernández 1991). This means that they are dynamic as the user's viewpoint is considered in each relationship, therefore an ability to model which Features of Interest (FOI) are in view is required to ensure only visible items are referenced.

7.3.2 Visibility Modelling

Most Geographic Information Systems (GIS) offer the functionality to carry out visibility modelling (De Smith *et al.* 2007), with a catalogue of research including siting radio masts (De Floriani *et al.* 1994a), locating the most scenic or most hidden routes (Stucky 1998), landscape planning (Fisher 1996), as a weapon surrogate in military exercises (Baer *et al.* 2005), and in examining spatial openness in built environments (Fisher-Gewirtzman and Wagner 2003).

Studies in the urban landscape have tended to be based on isovists (Tandy 1967), using in particular Benedikt's (1979) interpretation and definitions. Essentially isovists describe the space which is visible from a vantage point considering the form of the built environment through the use of architectural plans which denote the building footprint and position. However this model ignores building height, the topography of the land surface, and the continuation of the lines of sight beyond the first intersection with a building footprint. Therefore isovists depict lines which when traversed from the vantage point offer a continuous view of the target, and disregard more distant features.

Recently 3D isovists (Morello and Ratti 2009) and visual exposure models (Llobera 2003, Bartie *et al.* 2010) using DSMs built from LiDAR sources have been introduced for urban visibility modelling. These DSMs include building and topographical form and may be used to determine how much of a feature can be viewed from the surrounding space, enabling the

creation of surfaces to show in which direction an observer would need to move to view the target more, or less, clearly. These techniques can be used to find visual corridors, or visual ridges, and form a useful basis for considering feature visibility in the context of LBS.

The urban visual exposure model calculates the vertical extents visible for each building cell of the DSM, by calculating the lowest visible point on the façade from the intersection of foreground objects, as shown in Figure 7.1. From this the visible façade area, and the percentage of a feature on the skyline may be deduced, along with other metrics.

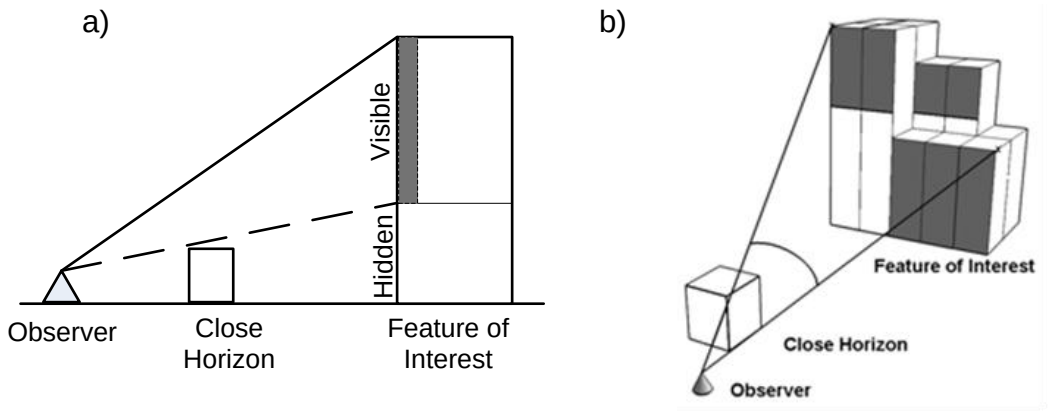


Figure 7.1: Visual Exposure in Urban Environments

Once the model is able to determine which features are visible, it is possible to then relate these to construct the positional part of a referring expression. To translate the positional information into the user's frame of reference requires an egocentric projective spatial model as discussed in the next section.

7.3.3 Egocentric Spatial Model

Natural language terminology relates space according to the observer's frame of reference, or that of other features in view. This involves turning the relationships from metric space into projective space using terms such as 'left of', 'right of', ' before', 'after' and 'between' as presented by the 5-intersection model (Clementini and Billen 2006). In addition a quaternary projective relation model is required to include terms for 'above', 'below', and 'coplanar' (Billen and Clementini 2006).

Figure 7.2 shows a combined model which uses these projective models for the space in front of an observer, and a simplified intrinsic frame of reference for items behind the user (Bartie *et al.* forthcoming). For two items to be considered opposite one another reference is inferred to a common central space, or object. For example, "the library is opposite the park" determines that from some viewing point both the library and park are visible and facing towards each other

from a common central region. The following section explores these concepts of the inter-relationship between visibility and projective relations with respect to the "opposite" relation.

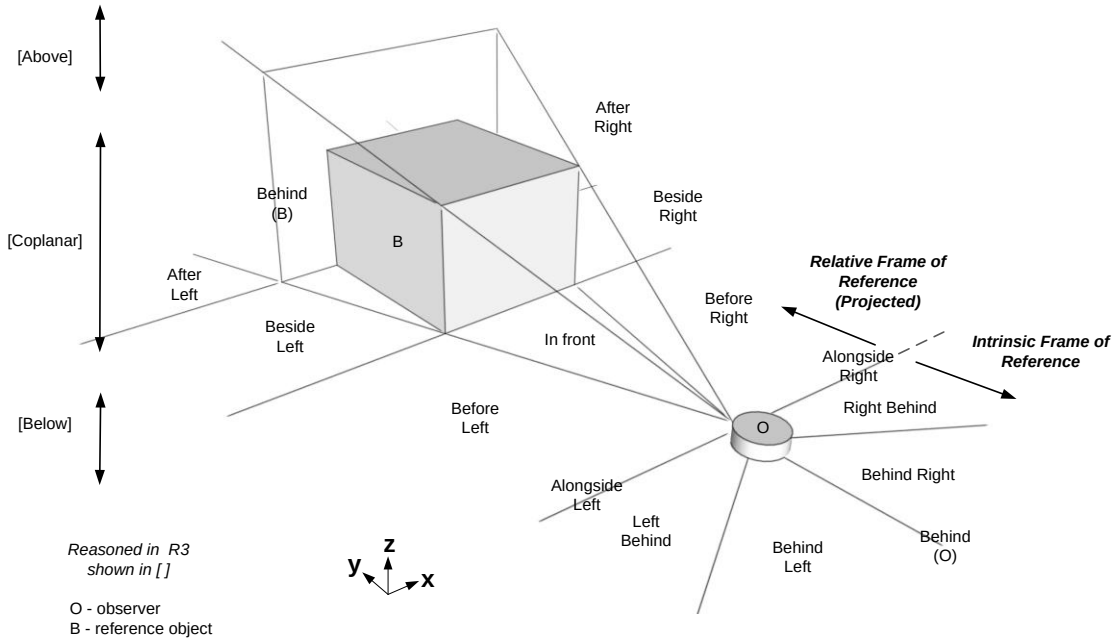


Figure 7.2: A combined model of space using relative and intrinsic frames of reference

7.4 The 'Opposite' relation

The spatial relation "opposite" is defined in Merriam-Webster's dictionary as, "*set over against something that is at the other end or side of an intervening line or space"* (Merriam-Webster 2010). Hence the interrelated visibility of three related physical entities are required, such as the library, park and street of the previous example. For this reason visual exposure modelling is required to report which objects may be used in the relation. Two cases are examined here, firstly where the entity is represented as a one dimensional feature, such as a road or river, and secondly where it is represented as a two dimensional region, such as a park.

7.4.1 One Dimensional Common Feature

The relationship conveyed in the phrase "the library is opposite the park", can be broken down into "the library is *left* of the road" and "the park is *right of* the road" from a viewpoint on the road. However this does not signify that the two objects are "opposite" each other unless both are perceived to occur at similar positions along the road.

Consider Figure 7.3 which shows a situation where a number of buildings surround a park, and the observer is located on a road at Point 1. From this viewpoint the observer is able to see Buildings A,B,C, and the upper part of D above hedges, but not Buildings E, F and G which are

out of view and therefore not included in any referring expressions. When the observer faces Building C, it is valid to report that both Buildings C and B are 'opposite' the park. Here the term "with reference to the road" is left out but inferred, and the phrase is equivalent to "Building C is on the other side of the road from the park". However it would not be appropriate to define Building A as "opposite the park", as the two features do not share a similar location along the linear road feature, yet the phrase "opposite side of the road" is still true. The segments of the road labelled A1-A2, B1-B2 indicate the start and end of the entity along the road's length, computed at the first and last intersections of a line perpendicular to the direction of travel with the entity. To satisfy the "opposite" condition the following must apply:

- the observer must be able to view both entities from a single point (e.g. C and Park)
- the entities must occur at overlapping sections of the linear entity (e.g. C1-C2 and P1-P2)
- the observer must also be able to view the common linear entity in the overlap region (e.g. road)
- the entities must occupy *beside left/right* space when viewed from the overlapping region.

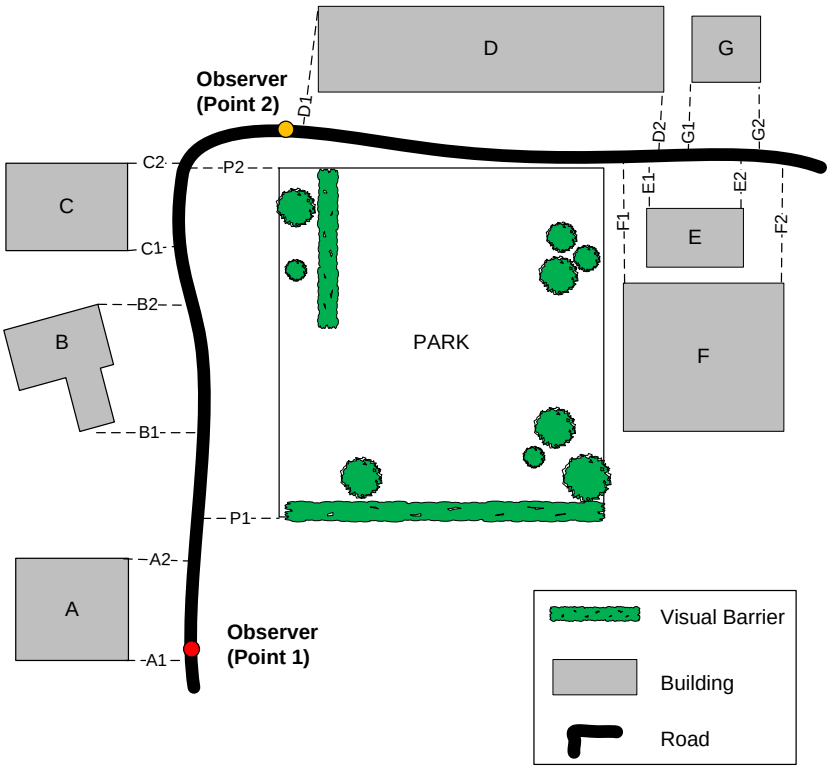


Figure 7.3: The 'Opposite' Case for with a Common Linear Feature

Following these rules Table 7.1 may be generated, indicating that B and C are opposite the park when viewed from Point 1. Although part of D is visible above hedges, the section of roadway between the building and the park is out of view behind bushes, rendering the use of "opposite" as less appropriate to assist the user's visual search from the current location.

However the term could be used if instructions considered the case as the user approaches, such as "when you get to *Point 2* you'll see D *opposite* the park".

When multiple features satisfy the 'opposite' relation further consideration is necessary to establish which would form the most suitable candidate. So from Point 2 buildings E, F and G come into view and may be considered as candidates for describing the location of building D. A function is required to establish which of these is most suitable, requiring knowledge of the candidates and consideration of the overlap extent. In this case the overlap between D and E is minimal, and although F has a larger overlap it would still make more sense to describe D as "opposite the Park", as these features share the greatest overlap and the Park is a very recognisable feature.

Table 7.1: Calculating the Opposite relation for Entities and the Park from Observer Point 1

Entity	Visible	Overlap	View Common Overlap	Beside Left/Right	Result
A	True	False	True	True	False
B	True	True	True	True	True
C	True	True	True	True	True
D	True	True	False	True	False
E	False	False	False	False	False
F	False	False	False	False	False
G	False	False	False	True	False

As a further example, when the observer is at Point 2 looking for building G then the park can no longer be considered opposite, however either building E or F could be used. Factors including the saliency of the building, its visibility, distance from the target, and the number of items between each should be considered. Assuming the visibility of both E and F were high (clear views) then E would form the most logical choice as it is the first item viewed from the roadside and most readily identifiable. However if F was a visually prominent landmark, such as a church, then it would take precedence despite being further from the target as its saliency allows it to form a more useful descriptor.

Saliency is a measure of the prominence of a feature in the neighbourhood, and there are methods to quantify such distinctiveness (Raubal and Winter 2002, Elias 2003b). Typically factors including visual appearance and semantic interest are considered by comparing items in the neighbourhood to establish the most easily recognisable and rare features. It is of particular importance in choosing candidates for forming referring expressions, as when targeting a building by describing it as opposite a 'tree' it may be logically true, but worthless if the entire street is filled with trees and all houses are opposite a tree. Therefore when constructing a referring expression the number of other entities which share a similar relationships need to be considered, to minimise the confusion caused by the statement.

Fuzzy classes may be used to establish the most attractive entity in an 'opposite' relation, by considering all alternatives and awarding class memberships between 0 and 1 according to a number of factors. The weighting between factors may be adjusted according to the current task, for example car drivers may favour number of items between as scanning opportunities are more limited while driving, whereas pedestrians may favour saliency as they have more freedom to view the surroundings and wish to locate the most prominent landmarks in a wider field of view.

$$\text{Most suitable entity} = f(V,S,N,D,O) \qquad \textit{equation 7.1}$$

where: V – visibility (degree of visibility of all items from a single observation point)
S – saliency (prominent, minimise confusability)
N – number of items between (measure of separation by entity count)
D – distance apart (close items preferred)
O – degree of overlap

A slightly modified set of rules are necessary when considering two dimensional common features as discussed next.

7.4.2 Two Dimensional Common Features

For linear common features the entity overlaps, used to identify whether features are opposite, were determined by considering the first and last intersections of a line perpendicular to the linear feature with each entity (as shown in Figure 7.3, e.g A1-A2). In cases where the common feature is a region, an alternative rule is required to determine overlap, and consequently opposition.

When the observer occupies a space inside the common region, for example standing in a square, then two features may be described as opposite one another by considering the observer as a central point with a feature occupying the *in front* space, and one in *behind* space. As an example, if the observer looks towards feature A as shown in Figure 7.4(i), then B can be classed as in the opposite direction, according to the in front/ behind relationship outlined in Figure 7.2. However if the observer is outside of the common region then the relationship may be calculated according to a division of space based on the Orientated Minimum Bounding Box (OMBB), as shown in Figure 7.4(ii). In this case the OMBB is drawn around the common region, and the centre point determined. Lines are extrapolated from the centre point to the corner and edge midway points of the bounding box, creating 8 triangular zones. For any two entities to be considered opposite each other with respect to the square they must occupy zones whose sum adds up to ten, according to the number system shown. Therefore no matter where the observer is located the relationship of A, B1 and the square would be classed as 'opposite', assuming all

entities were visible. Entities occupying a neighbouring zone are also considered to be ‘opposite’, so that A, B2 and A, B3 would also be described as sharing an ‘opposite’ relation with respect to the square. This works for all common region shapes (e.g. lakes), as the algorithm is based on the orientated bounding box.

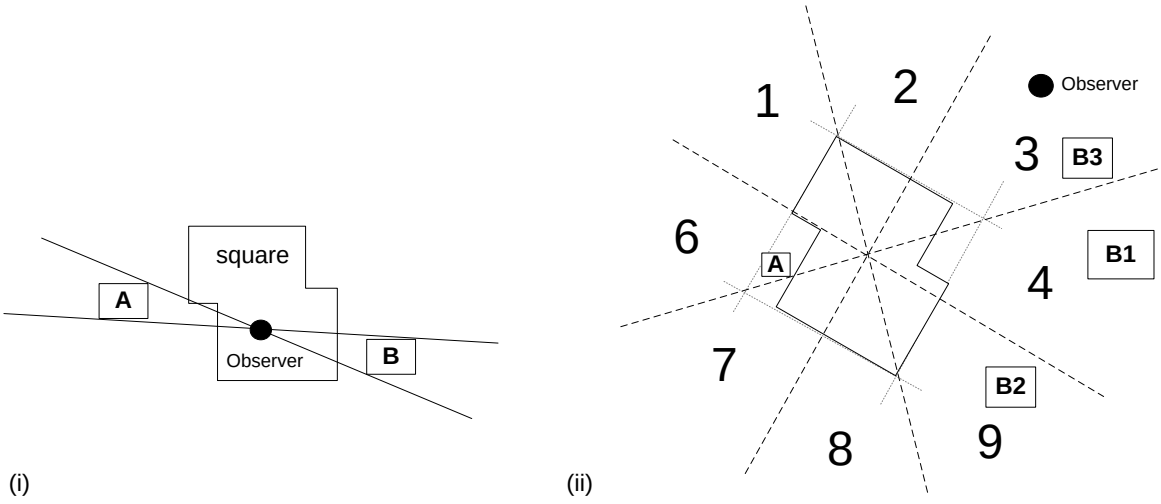


Figure 7.4: Using the term “Opposite” across a region

As before other factors are considered when it is necessary to determine the most suitable candidate if more than one feature satisfies the relationship, such as object sizes, distances apart, and the number of confusable similar entities in view.

7.4.3 Additional Considerations for Usage

The concepts outlined so far have failed to include a measure of the appropriateness of the inclusion of the term with respect to the observer’s viewing distance, and object sizes. Instead they have looked to determine the most suitable candidate for the relation. However at greater viewing distances, and for smaller items, it may be harder for the observer to judge when two entities share an opposite relation and it may be necessary to restrict the inclusion of the term in a referring expression according to the percentage of the field of view occupied by the items. This approach accommodates entity scale, such that a house may be described as opposite a bakery from only close range, while a park may be described as opposite the hills from greater distances.

Additionally when a referring expression is used in a more general description of a region, and not to identify a particular target, consideration must be given to the ordering of features in the relationship. Jackendorf (1992) makes the observation that not all relationships are symmetrical, and that “the house is next to the bike” makes less sense than “the bike is next to

the house". When considering the "opposite" relation it makes more sense to use the most salient feature as the reference object, such as "turn right after you see a hut opposite a lake", whereby the viewer's attention should be more naturally drawn to the lake, and the decision point confirmed once the hut has been located.

7.5 Conclusions and Future Work

When constructing descriptions of a feature it is useful to include spatial prepositions to guide the user's attention. This is particularly relevant when forming referring expressions for use in speech based LBSs when directing the user's attention to items in view while exploring a city. The case of spatially 'opposite' an entity has been considered in this chapter raising a number of observations about how it may be determined and constructed from GIS datasets for use in LBS applications.

The research has shown that to ensure meaningful descriptions it is necessary to determine whether features are visible to the user by calculating their visual exposure and establish a common reference entity to define their relationship. Here one dimensional linear, and two dimensional regional features have been examined for this purpose. Three dimensional entities have not yet been explored, but could be included in future work. They may be of particular use when constructing descriptions for features on the side of a building, such as "the window opposite the balcony", to limit the vertical scan region.

When a number of possible candidates are found it is necessary to select the most useful by determining its saliency and recognisability, to assist in guiding the user's attention and minimise risk of target confusion. Factors including its visibility, distance, and the number of other similar items in the scene are considered to ascertain the best candidate. Future work should examine the weighting of these inputs, to determine the most suitable values for particular tasks.

In a wider context such definitions are useful to provide a geosemantic bridge to a future semantic web and publish the "opposite" relation in a manner that it can be used automatically. Two approaches for achieving this include, firstly, within a description of the concept "opposite" in the ontology a callout can be inserted, which links to an external web service, or secondly the nature of that function could be defined in something akin to an ontology, which is here called an epistemology.

Given the first approach of calling out to an external service, a pointer in the ontology would direct the user, or software agent, to a web service that computed the opposite relation. The service would include a semantic wrapper (e.g. in OWL-S, WSDL, or WSML), defining what the service does and the type of inputs needed and outputs produced by the service, see for

example Bruin *et al.* (2008). The limitation of this approach is that the algorithm for the "opposite" relation is semantically a black box and cannot be inspected or reasoned with.

The second approach for publishing a semantically well defined "opposite" relation, involves describing the algorithm in full, that is, how it works rather than what it does. Gruber (1993) classically defines an ontology from an AI perspective as a specification of a conceptualisation. From a similar perspective, an epistemology might then be defined as a specification of functions that contribute to the concepts in the conceptualisation, that is, a specification of how we come to know what we know. This epistemology might be expressed in some kind of pseudo code or declarative language, such that any software might be able to run it, and is linked to an ontology that describes other aspects of the concepts used. The advantage of this approach over the former is that the functions are no longer black boxes, which might support new kinds of reasoning such as the comparison of epistemologies.

The models presented should be developed further through user trials, and may then be adopted as part of a wider set of defined spatial relations for use in urban LBS, and the geosemantic web.

Chapter 8: A Combined GIS and Stereo Vision approach to Identifying Building Pixels in Images to Determine Appropriate Colour Terms

8.1 Summary

Colour information is a useful attribute to include in a building's description to assist the listener in identifying the intended target. Often this information is only available as image data, and not readily accessible for use in constructing referring expressions for verbal communication. The method presented uses a building polygon GIS layer in conjunction with street level captured imagery to provide a method to automatically filter foreground objects and select pixels which correspond to building façades. These selected pixels are then used to define the most appropriate colour term for the building, and corresponding fuzzy colour term histogram. The technique uses a single camera capturing images at a high frame rate, with the baseline distance between frames calculated from a GPS speed log. The expected distance from the camera to the building is measured from the GIS layer and refined from the calculated depth map, after which building pixels are selected. In addition significant foreground planar surfaces between the known road edge and building façade are identified as possible boundary walls and hedges. The output is a dataset of the most appropriate colour terms for both the building and boundary walls. Initial trials demonstrate the usefulness of the technique in automatically capturing colour terms for buildings in urban regions.

8.2 Introduction

When talking about a place people like to include descriptive words to conjure up a pictorial representation in the listener's imagination. Such feature descriptions are also often included in way finding instructions, such as the details of a building façade material or colour, as in "We're the red brick house with the long white fence.". Capturing this level of detail has received a lot of attention in recent years, including initiatives such as Google's Street View (Vincent 2007) and Microsoft's Street Slide (Kopf *et al.* 2010). However texture information is presented as imagery and not transferred into values suitable for use in cartographic symbology, or for inclusion in way-finding feature descriptions. This chapter presents a method whereby depth mapping is used to automatically filter foreground objects from images, allowing the automatic extraction of colour terms to describe buildings, so that a database of building colours may be created.

Object descriptions are known in natural language research as "referring expressions" and are used, for example, to draw someone's attention to a particular building in a cityscape (Dale and Reiter 1995). They include visual clues which the speaker considers to be useful aids for the listener to determine which item in view is the intended target. The most useful terms are those which the listener can identify quickly, and limit the number of candidates rapidly without leading to any confusion. In some ways the process of creating a referring expression is similar to determining landmark saliency, which focuses on ways to measure the prominence or distinctiveness of a building according to a number of factors, including visual, semantic, or structural attraction (Sorrows and Hirtle 1999). There are two main methods for extracting landmark candidates, by assigning a saliency score based on various attributes. Elias (Elias 2003b) uses characteristics such as the building area, number of corners, density of buildings in the district, orientation to north and so on, while an alternative definition for saliency measurement was proposed by Raubal and Winter (2002), later updated with Nothegger (Nothegger *et al.* 2004), which scores buildings according to Sorrows and Hirtle's (1999) visual, semantic, and structural characteristics. The visual factors include façade area, shape, colour, and visibility, translating well to the egocentric projective view experienced by street observers. These visual variables closely reflect Bertin's (1983) set of seven visual variables (position, orientation, size, colour, value, texture, and form) which should be considered when displaying graphical information. While traditional GIS datasets store position, orientation, and planimetric size of buildings, they fail to show information relating to building height, colour or texture. A challenge exists therefore in how this information may be sourced and made accessible in a format suitable for use in constructing referring expressions. While LiDAR now offers a viable solution for capturing building height and form (Palmer and Shan 2002, Rottensteiner and Briese 2002), colour and texture details are either unavailable or stored in an inaccessible form, such as street level images.

Oblique aerial imagery could provide a source for the missing colour information, offering more detail on the sides of buildings than can be obtained from traditional overhead aerial imagery. However while the textures can be directly mapped on to the surfaces of building models (Lensch *et al.* 2000, Frueh *et al.* 2004), foreground objects such as cars and trees are indistinguishable and are incorrectly included in the building façades. New techniques are being developed which attempt to remove the unwanted foreground elements (Forsyth *et al.* 2008), or fill in the background using images from alternative angles (Yi-Leh *et al.* 2010). A sensor fusion approach has also been successful where laser ranging equipment is used in conjunction with cameras to collect textures (El-Hakim *et al.* 1998, Pylvänäinen *et al.* 2010, Wang and You 2010). While these techniques look to offer the solution to produce clean façade textures in the near future, the timescale is unknown and current coverage is sparse.

In the meantime the research presented here offers a way to capture building colour details from street level using low-tech equipment available to most communities, automatically excluding foreground objects allowing the remaining building pixels to be classified with an appropriate colour term. The challenges are how to automatically identify which pixels in an image correspond to a building, classify the colours in those pixels using the most appropriate colour term, and then associate those values with the correct building polygon on a map.

There are a range of applications which would benefit from having access to building colour details, including the ability to generate more natural way-finding instructions (Raubal and Winter 2002, Nothegger *et al.* 2004), which would be especially beneficial for children (Helvaciog lu and Olguntürk 2009) and people with learning difficulties (ODPM 2006). The ability to create more descriptive navigational instructions which would be indistinguishable from those generated by a human has been referred to as the 'spatial Turing test' (Winter and Wu 2008). In addition colour can be used in forming referring expressions (Dale and Reiter 1995), which are particularly useful for describing objects in 'vista space' (Montello 1993), that is the region currently visible to an observer. Being able to verbalize colour information will be an important component in the future of speech interfaces whereby a user operating both hands-free and eyes-free may request information on a building in view, selected by its description (Michelis *et al.* 2008). A Location Based Service (LBS), such as a virtual city guide application (Bartie and Mackaness 2006), may direct the user's attention to a specific building using a narrative which singles it out in the current view, simulating a natural language description. Emergency services would also benefit from access to a building colour database when attempting to locate people based on a description of their surroundings (Le Yaouanc *et al.* 2010).

The chapter is arranged as follows: Section 8.3 discusses two appropriate computer vision techniques which can be used to calculate pixel depth values, enabling the selection of pixels in the image at distances which correspond to that of the designated target building as explained in Section 8.4. These pixel values may be translated into relevant colour terms using a fuzzy set approach as discussed in Section 8.5. A trial of the proposed method is demonstrated in Section 8.6.

8.3 Stereo Depth Mapping for Façade Colour Retrieval

There are a number of methods which may be used to recover depth information from images, including the Structure from Motion approach (Koenderink and Van Doorn 1991, Sturm and Triggs 1996), and the Stereo Vision method (Lucas and Kanade 1981, Kanade *et al.* 1996, Birchfield and Tomasi 1999, Scharstein and Szeliski 2002).

The requirement here being to determine the distance from the camera to the real world object represented in each image pixel, such that when combined with a GIS building layer those pixels corresponding to a designated target at a known distance may be retrieved, while foreground objects are excluded. Once the depths per pixel have been recovered then those within the expected range are selected, as outlined in Figure 8.1.

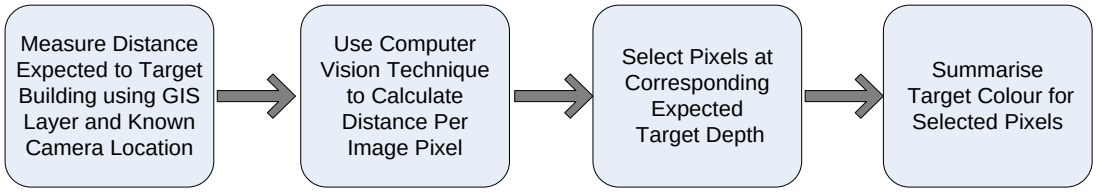


Figure 8.1: Overview of the Process to Select Target Pixels from Captured Image

8.3.1 Structure from Motion Approach

Through combining many views of the same real world object from different distances and angles, it is possible to reconstruct the object's structure. This is achieved through a processing pipeline which begins with matching pixels in at least three images. Pixel matching is an automated procedure, whereby points with high contrast gradients are found, such as SURF-based features (Bay *et al.* 2008). These features are tracked between images, known as correspondences, and the 3D structure of the real world object is recovered along with camera pose estimates.

The following images (Figure 8.2) were rendered from a model built using this Structure from Motion approach, whereby 26 images collected while walking in an arc around a property were processed using PhotoScan software (AgiSoft 2010).

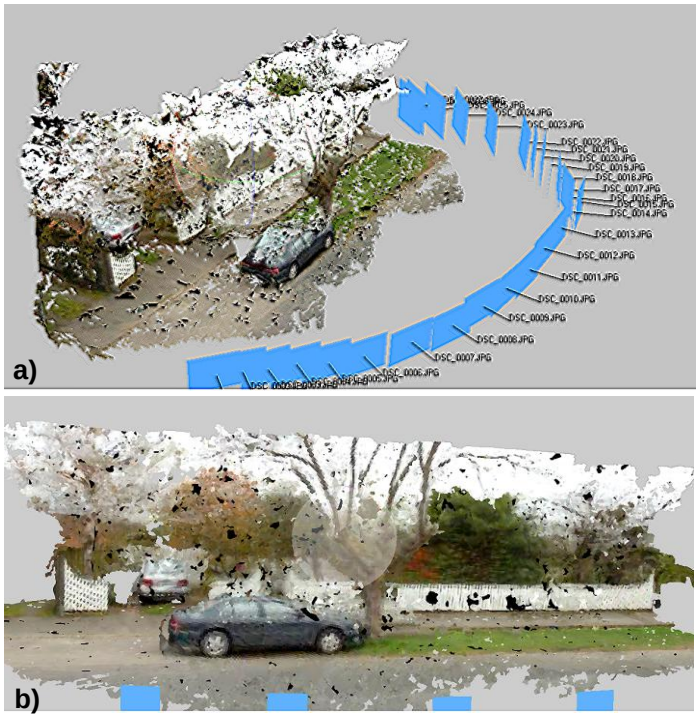


Figure 8.2 Images Rendered from a Model Constructed using Structure from Motion Approach

There are a number of considerations when using this technique in urban regions. Firstly a large number of images are required from a variety of viewing angles; a linear set of images from a single drive-by will not be sufficient to produce a detailed model. In addition to produce good results the source images need to be captured at high spatial density, meaning a new image is required each metre or so. Furthermore a robust solution requires that features need to be visible in at least three images captured from different locations, which can be rather difficult to achieve in confined streets, through foreground vegetation, or when there are many moving pedestrians and cars. Additionally processing times are fairly high, as feature matching, bundle adjustment (Triggs *et al.* 2000), and geometry reconstruction are computationally intensive tasks. Finally low texture regions lack features for the matching process, resulting in poor depth estimates.

There have been a number of attempts to overcome these shortcomings, such as sourcing images from popular image sharing websites such as FlickR to build community volunteered virtual models (Snavely *et al.* 2006), and improvements in depth reconstruction by imposing model restrictions such as enforcing planar surfaces (Micusik and Kosecka 2009). However currently this approach is best suited to capturing information in the more open urban spaces where a large number of images may be captured from a multitude of angles. A good example of this exists for Cathedral Square in New Zealand, where Photosynth user Redpaw (RedPaw 2008) used 330 images to construct a scene. A section of this which corresponds to the Christchurch Cathedral is shown in Figure 8.3, depicting (a) the point cloud and (b) an overview map

generated from the process. To assist the reader a sample image (c) from the captured image set, and an aerial image from the location as seen in Google Earth ™ (d) are included. The points in the data cloud may be georeferenced and rectified to fit the corresponding GIS layer, and as the colour details from the original image pixels are maintained it is therefore possible to retrieve colour information for any part of the building where a correspondence exists.

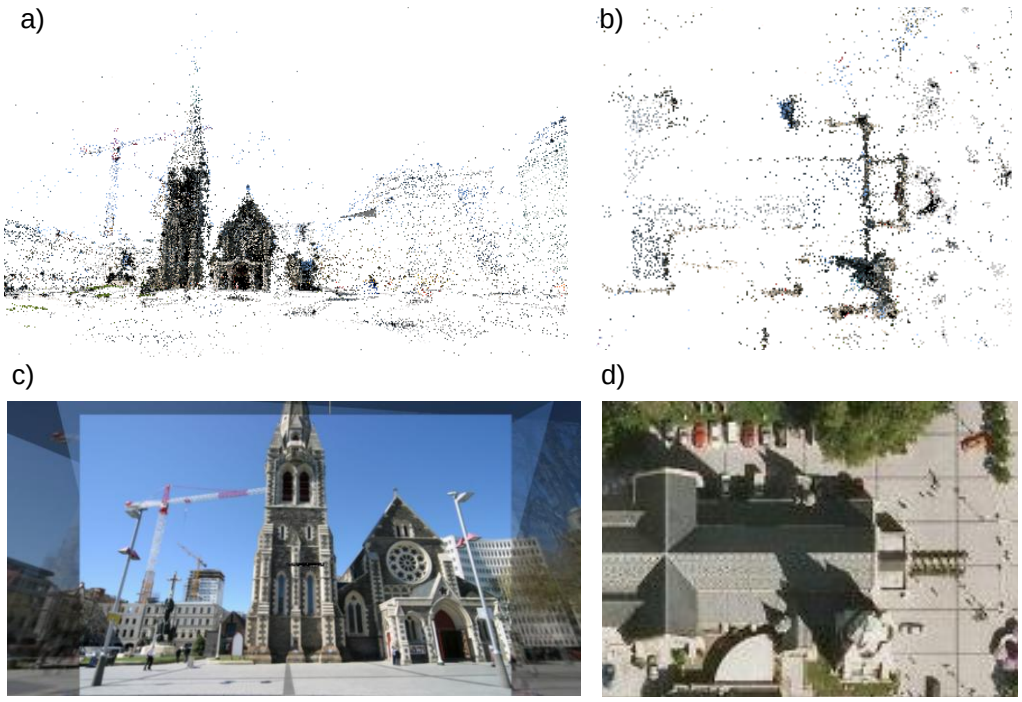


Figure 8.3: a) 3D Point Cloud b) Map View of Point Cloud c) Sample Image of Cathedral Square, Christchurch Captured by RedPaw (2008) d) Equivalent Region View in Google Earth (TM)

This technique shows promise for mapping open expanses, but the high number of overlapping images required to produce high density point clouds restricts its use in more confined and cluttered spaces. Therefore a less demanding approach which requires only a single stereo image pair, taken a known distance apart, is considered next.

8.3.2 Stereo Baseline Vision

Stereo baseline vision works by comparing two images taken parallel to each other but a short distance apart, known as the baseline distance. The technique requires the same feature to be identified in both images, so that the horizontal disparity may be measured. Larger disparities indicate objects are closer to the camera, as a result of distance parallax.

Computer vision techniques are used to identify stereo correspondences by locating interesting features (e.g. corners, edges) in one image, for example using Harris (Harris and Stephens 1988), Förstner (Förstner and Gülch 1987), or SURF (Bay *et al.* 2008) definitions, and search the paired image for the most similar matching template. The automatic process produces

a disparity map, which can be transformed into a depth map with knowledge of the camera's focal length, pixel size, and the baseline distance between the image pair (Bradski and Kaehler 2008).

In order to collect a sample dataset to trial colour extraction without foreground objects in a suburban region, a simple stereo rig was built from two low cost webcams set a baseline distance of 60cm apart. It is necessary to perform a one off calibration of the stereo camera setup to calculate various intrinsic and extrinsic details, allowing lens distortions to be removed from future images. This was done by capturing multiple views of a flat chessboard pattern using Emgu.CV (Canming 2010), a C# implementation of OpenCV (Intel 2010). Initial static trials of the rig showed that depth could be recovered successfully up to a distance of around 30 metres; however the webcams suffered from various distortions when moving at 50km/h due to their shutter design, rendering the images unsuitable for depth mapping.

Instead a single high quality video camera with optical stabilization was used to capture 25 images per second with a shutter speed of $1/500^{th}$ second to ensure sharp images without motion blur. The camera was fixed perpendicular to the direction of travel, and a GPS device was used to log speed, orientation, and location at 1Hz. The advantage of this setup is that it is extremely simple to implement, requiring only intrinsic details for a single lens, and uses technology available to a wide audience, potentially allowing it to be implemented on public transport vehicles such that urban colour datasets could be regularly updated.

To solve for depth using the stereo vision approach requires an input of the baseline distance between image capture locations. This can be calculated by measuring the ground distance travelled between frames, which for a car travelling at 50km/h would be a baseline distance of 56cm, as shown in example A of Table 8-1. Rather than using GPS location information and measuring the distance between shutter releases, the relative offset between frames may be calculated more accurately using GPS speed. The main difference is that GPS speed is considered to be accurate to within 0.2m/s (0.72km/h) (Witte and Wilson 2004), and is calculated using Doppler shift making it more robust in multipath environments. At a frame rate of 25 images per second the speed accuracy would constitute a baseline discrepancy in the region of 0.8cm, resulting in negligible depth inaccuracies. This can be observed in Example B of Table 8-1, where an increase in camera speed of 0.2m/s results in a baseline distance 0.8cm greater than shown in Example A.

Table 8-1: Example of how to calculate baseline distance

Example	**Frame Rate per Second**	**Speed (km/h)**	**Speed (m/s)**	**Baseline Distance (m)**
A	25	50.00	13.89	13.89 / 25 = 0.5556
B	25	50.72	14.09	14.09 / 25 = 0.5636

To ensure the most accurate baseline estimate for any frame pair, the speed was interpolated from the surrounding GPS measurements, as shown in Figure 8.4. Here a captured frame (801) falls between two GPS readings, resulting in an interpolated speed, from which the baseline distance may be calculated (ground distance between frames 800 and 801) so that disparity information may be translated into real world depth units (metres).

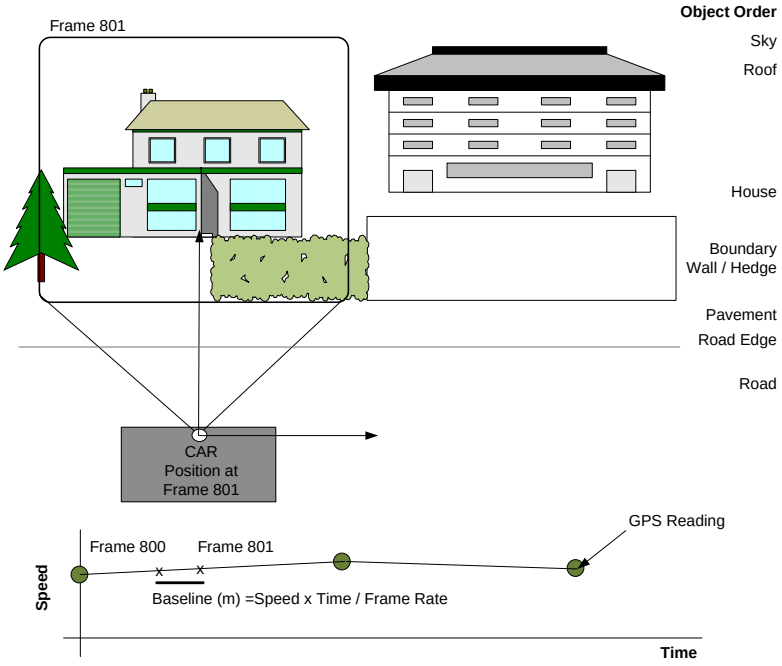


Figure 8.4: Calculating the Baseline and Expected Disparity

The relationship between disparity and depth is inversely proportional, and being non-linear high depth resolution is only available for objects near the camera. The equation is given as,

$$d = f\ B\ p\ /\ Z \qquad \textit{Equation 8.1}$$

where d is the disparity expected between features in the image, f is the camera's focal length, B the baseline distance moved between frames, p the pixels per centimetre on the camera's sensor, and Z is the distance from the camera to the house (Bradski and Kaehler 2008).

An example of this relationship for a camera travelling at 50km/h is shown in Figure 8.5, plotting disparity (right-side y-axis) against world object distance (x-axis). In addition the left-side y-axis displays the difference in pixel disparity which would be observed for a 1 metre change in object distance, considered as a reasonable margin of error for limiting the incorrect inclusion of any cars parked in front of buildings. As an example an object at a distance of 23 metres from the camera would have a disparity of 24 pixels, while an object 24 metres away

would have a 23 pixel disparity. This is the limit at which a 1 metre change in distance can be measured at the pixel level for this camera at a speed of 50km/h, essentially defining the maximum reliable working depth. Beyond this the recovery range can be increased by using a larger baseline accommodated here by using a 2 frame offset between left and right images, extending the 1 metre depth resolution limit to a distance of 32 metres. A 3 frame offset extends the search resolution to 40 metres from the camera, however greater offsets result in less similar foreground views, rendering it more difficult to find stereo correspondences for closer items (Scharstein and Szeliski 2002).

A slower moving car will result in a shorter baseline between images, dropping depth resolution for distant items. This can be overcome automatically by calculating the current forward speed for each frame and the expected distance to the target building from a GIS layer, and then choosing the lowest frame offset suitable for the required depth, thereby balancing the highest level of stereo correspondence matching and depth resolution. This approach offers both simplicity in data capture, and reduced computational requirements in data processing, with the flexibility to capture objects at a wide range of distances, and was therefore adopted for this research.

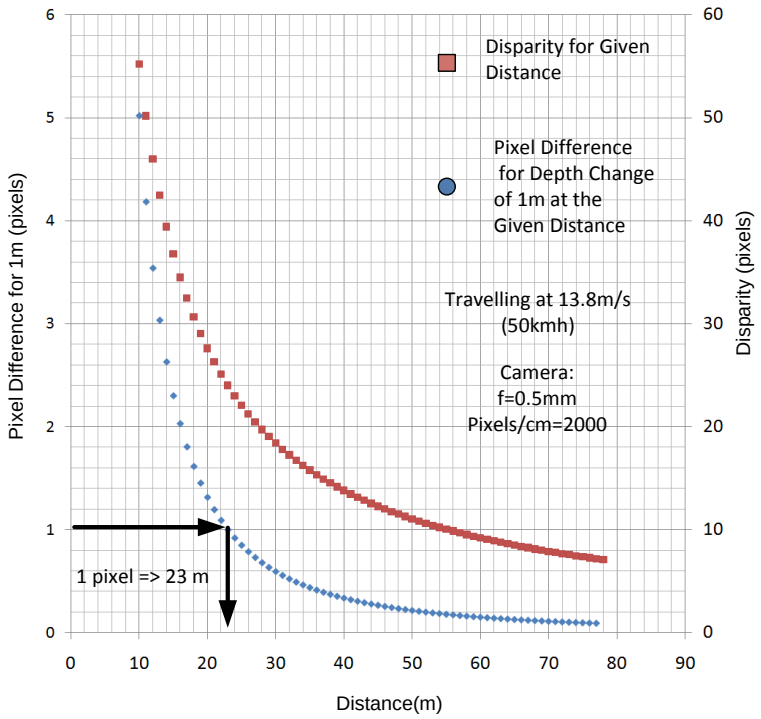


Figure 8.5: Example of the Relationship between Distance and Pixel Disparity

8.4 Processing Pipeline

The processing pipeline implemented is summarized in Figure 8.6, followed by explanations of the components. The process begins with the identification of a target building from a GIS

dataset, after which the nearest GPS data sample point is determined and the corresponding frame retrieved from the image stream. The distance from the camera to the target is measured from the GIS layer, and in conjunction with the camera speed at the time of image capture used to determine the optimum baseline distance required between the stereo image pair. In addition the baseline distance is used to transform the disparity map into a depth map (Section 8.3.2), refined using a depth frequency histogram approach (Section 8.4.2). The pixels at the corresponding target depth are selected, with those matching the current sky hue dropped from the selection (Section 8.4.4). If the pixel count remains above a specified threshold the image is considered suitable for inclusion in the classification, otherwise the next image in the sequence is processed. Before classifying a Retinex filter is applied (Section 8.4.5), and the image blurred and dilated slightly to remove pixel colour noise. The selected pixel colour values are saved as an array with the target building identification number, and the process repeated until ten locations are collected for each target building. In some cases it was impossible to collect enough good data for a target and so these were marked as irretrievable.

At this point in the processing pipeline two pixel groups were identified, one for the most likely house pixels and the other for any likely boundary wall (Section 8.4.3). The final stage of the process determines the most appropriate colour terms for each of these groups, as explained in Section 8.5.

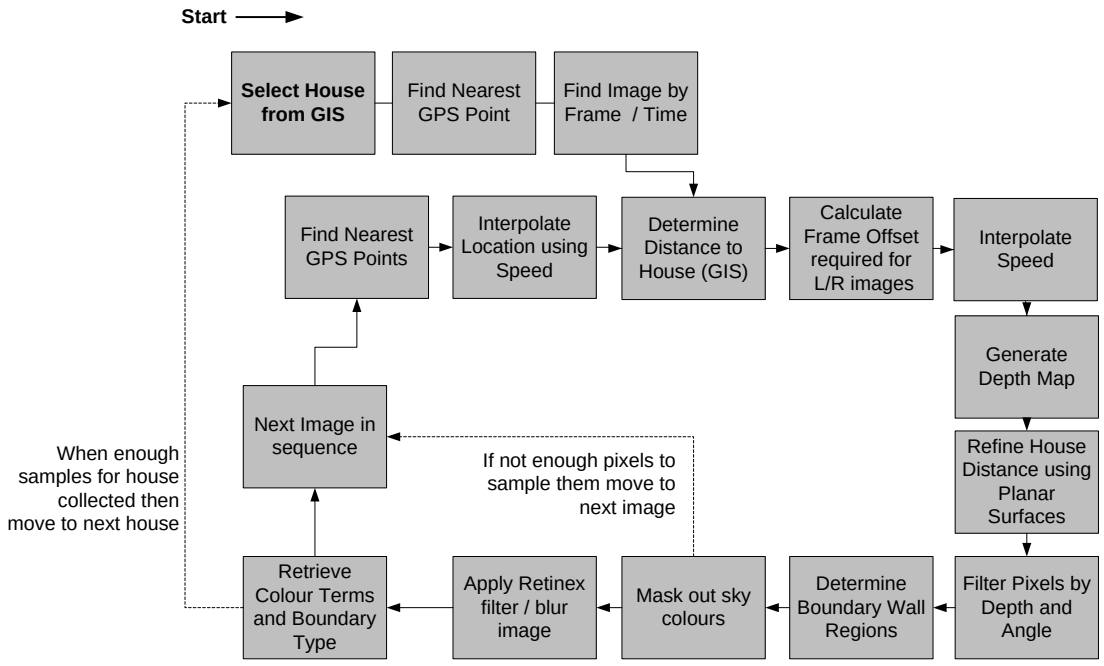


Figure 8.6: Detailed Processing Pipeline

8.4.1 Sampling Strategy

A sampling strategy was implemented to ensure that a number of images would be captured for each target, and should the building be obscured behind foreground objects that alternative views would be used. In all cases the first sample was set at the mid-point along the building's length from the road, with subsequent samples taken radiating outwards from this point offset a frame in either direction (i.e. +1 frame, then -1 frame, then +2 frames, -2 frames etc), forming a sampling sequence as shown in Figure 8.7.

The camera's position for each sample was interpolated from known GPS locations, weighted according to the acceleration or deceleration experienced at that time. It proved necessary to also estimate the approximate horizontal position of the building within the image frame, so that objects at similar depths to the side of the intended target, such as neighbouring houses, could be filtered out from the pixel selection process. This was achieved by projecting the building's edges into the image based on the known camera's field of view, which was calculated in the initial camera calibration stage and remained constant throughout data capture.

Each sample location was only considered valid if the number of pixels successfully placed within the expected house location was above a given threshold. For our trials this threshold was arbitrarily set to 100 pixels, deemed the minimum requirement for a fair reflection of the house colour. It was also considered that 10 successful sample locations, giving 1000 pixels in total, should be used before selecting the next target feature.

The sampling approach is depicted in Figure 8.7, whereby the locations A and B would be expected to give the highest pixel counts for each building, however on occasions where high fences or vegetation restricted the view the most successful samples were collected across driveways, such as at location C between houses. It is in these cases in particular that the estimation of the horizontal space occupied by the target is required, ensuring that only those pixels relating to the designated target are included in the colour summary.

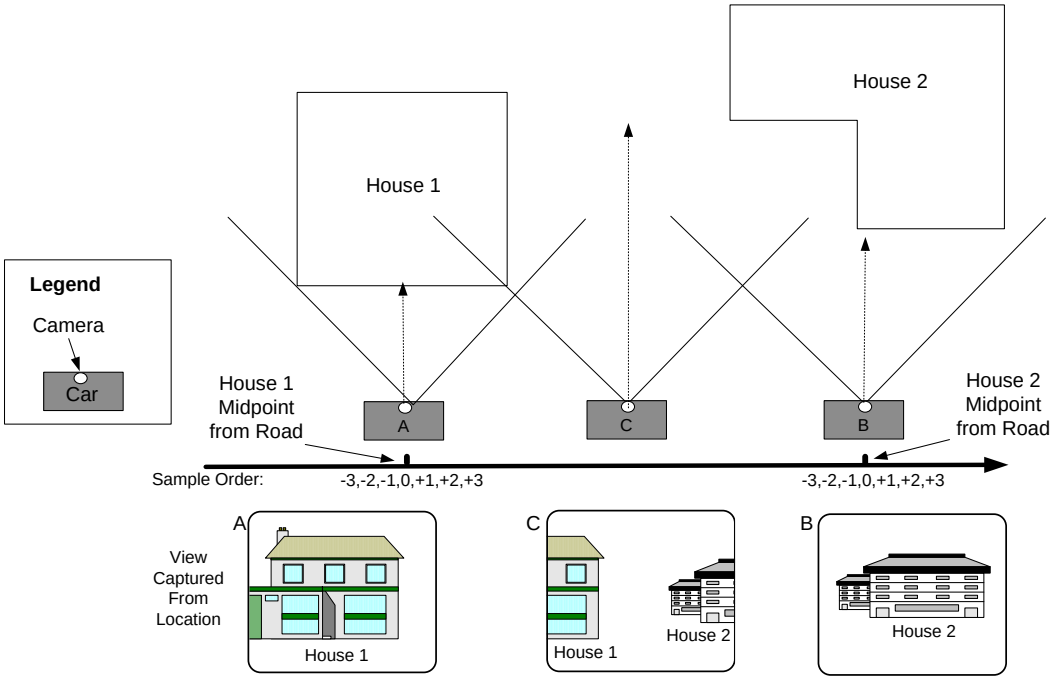


Figure 8.7: Estimating House Position in Image

As neither the GPS location nor GIS building layer inputs are error free additional steps were taken to improve system robustness.

8.4.2 Improving System Robustness

To accommodate system noise, from errors introduced in the GPS, GIS, or baseline calculations a further step was added to the processing pipeline. By summarizing the depth map vertically (against the x-axis) a frequency graph was generated which shows the most commonly occurring depths. Vertical planar surfaces result in high frequency depth counts, as a high proportion of the column exhibits the same depth value. Neighbouring high frequency values across the graph (y-axis) indicate a wide planar surface exists, such as a building or wall.

Figure 8.8 shows an example of the depth frequency histogram superimposed and aligned onto the corresponding image, with only significant values displayed after a high pass filter has been applied to remove low frequency noise. This depth summary was used to refine the expected depth value for the target building, thereby correcting small discrepancies in the input variables. In addition a depth tolerance of 1.5 metres was permitted in all cases, judged to be shallow enough to exclude cars parked in front of target building, yet giving some flexibility to the pixel selection process.

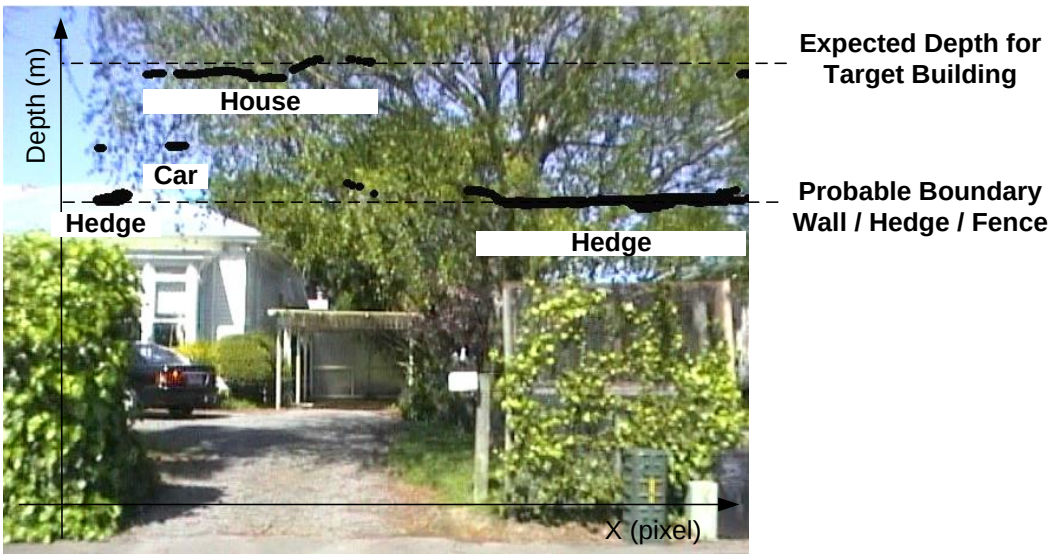


Figure 8.8: Using a Depth Frequency Histogram to Refine the Selection of Target Building Search

8.4.3 Identifying Property Boundary Type

In addition to searching for building colour information it was also possible to retrieve information on boundary fences and wall colours by searching in the lower part of the image for the existence of planar surfaces, occurring at a distance between that of the road edge and target building (see Figure 8.4). Planar surfaces were identified by passing a 3 by 3 kernel over the depth map to compare the depth gradients between cells, identifying regions of constant gradient. Figure 8.9a shows a theoretical example of this where a fence would exhibit linear gradients in both x and y directions, while vegetation would result in non-uniform gradients. By using a recursive function it was possible to identify significant planar candidates with similar gradient properties, merging results to define the larger planar surfaces, which in turn were filtered to leave only those vertical and facing the camera. Figure 8.9b shows the planar surfaces facing the camera which were recovered from the scene depicted in Figure 8.8.

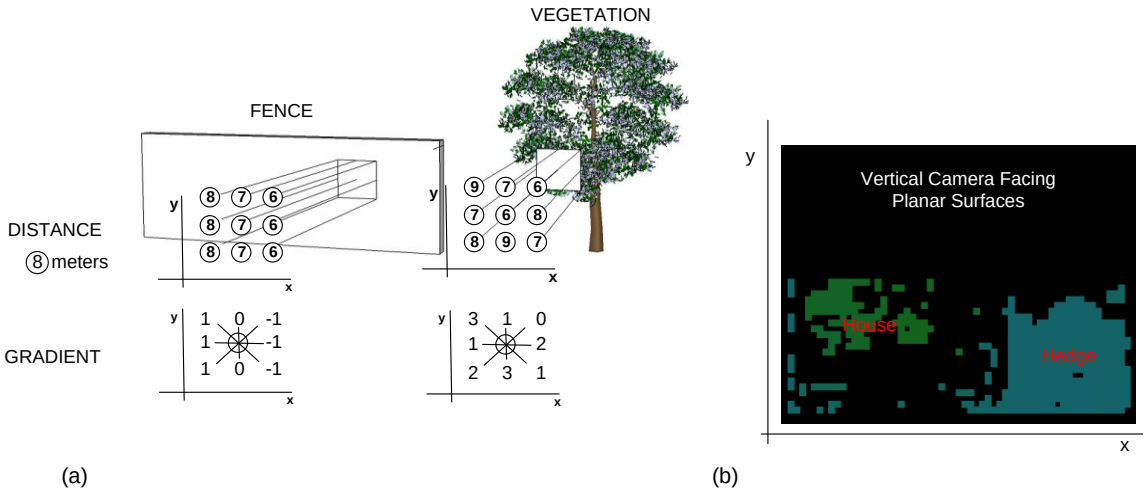


Figure 8.9: Detecting Planar and Non-Planar Surfaces using a 3 by 3 Kernel

The existence of any large flat surfaces at the property boundary could be identified by using the output from the depth frequency histogram (Figure 8.8) in conjunction with the existence of these planar regions, compared to the known road edge distances measured from the GIS road dataset. An example of the output from each stage of the process from the captured stereo views, to generating the disparity map, to selecting the house pixels, and discovering planar surfaces is shown in Figure 8.10

As well as discovering the boundary colour an attempt was made to determine the boundary material. Material recovery is recognized to be a difficult task (Hayman *et al.* 2004), and therefore to simplify matters the options were limited to a predetermined list of wall/fence, slat fence, or vegetation.

Hedges could be recognized due to the dense counts of Harris corners (Harris and Stephens 1988), while slat fences could be found by applying a Gaussian blur to the image before running the Canny edge detector (Canny 1986). The part of the image corresponding to the boundary was then scanned horizontally at a number of places tallying the number of intersections with detected Canny edges. Regions exhibiting a similar number of intersections across each horizontal scan, those with a low standard deviation, were deemed to be likely slat fence candidates, as shown in Figure 8.11. Other borders which did not satisfy either of these conditions were labelled as wall / fence.

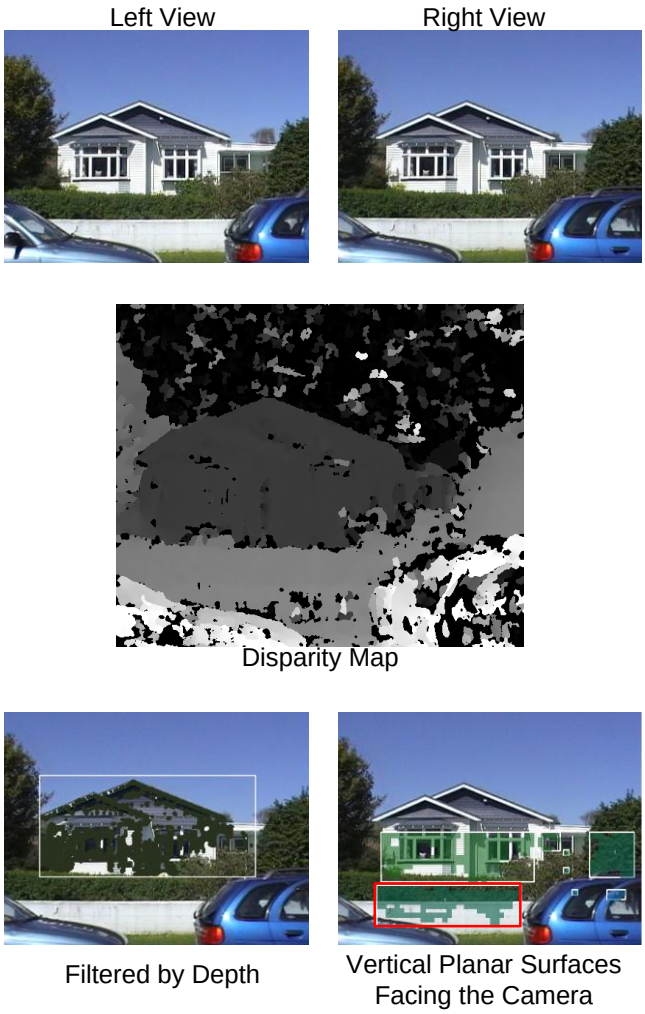


Figure 8.10: Stereo vision used to build a disparity map, and turned into a depth map used to identify building pixels. The most likely boundary wall/fence region highlighted in red.

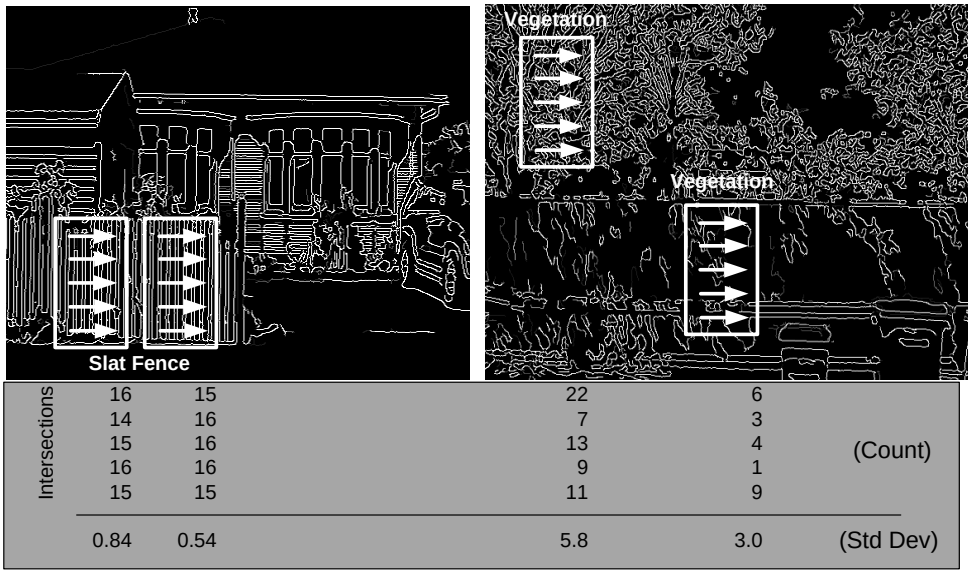


Figure 8.11: Slat Fence Detection using Canny Edges and Showing Intersection Counts and Standard Deviation

8.4.4 Windows

One of the issues encountered when automatically recovering building colour histograms is the presence of windows, which have no colour of their own but either reveal the interior, or reflect the surroundings. There are methods which can be used to automatically determine window locations based on gradient projection approaches (Recky and Leberl 2010), however these require the extent of the façade to be pre-defined in each image. A more simplistic approach is to filter the image for sky hues in HSL (i.e. Hue, Saturation, Lightness) colour space, thereby identifying those surfaces reflecting the sky. The pixels selected with similar colour may be removed from further consideration, leaving a smaller set of candidates with a higher likelihood of being part of the façade. It was found that on both bright sunny days and bright overcast days the procedure worked fairly well, although future research should look to implement more sophisticated procedures to identify windows under all lighting conditions.

Figure 8.12: Avoiding window regions from selection using sky hue filter

8.4.5 Shadows

The colour summary is also complicated by shadow regions, which are darker patches resulting from changes in lighting as a result of surrounding features (Finlayson *et al.* 2002b). A Retinex filter (Land 1977) may be applied to images to reduce the effect of illumination variation, as shown in Figure 8.13. Here an image is reduced to use the closest of only 11 colours before and after the Retinex filter is applied, showing an improvement in colour matching for regions in shadow (such as trees) after processing.

Although the Retinex process improves the colour classification, some illumination artefacts remain in the image (e.g. strong shadows show up as sky blue), and further shadow reduction techniques may prove beneficial (Scanlan *et al.* 1990, Finlayson *et al.* 2002a, Jianhong *et al.* 2010). The issue of strong shadows is reduced by limiting data capture to bright overcast days, when the lighting source is more diffuse.

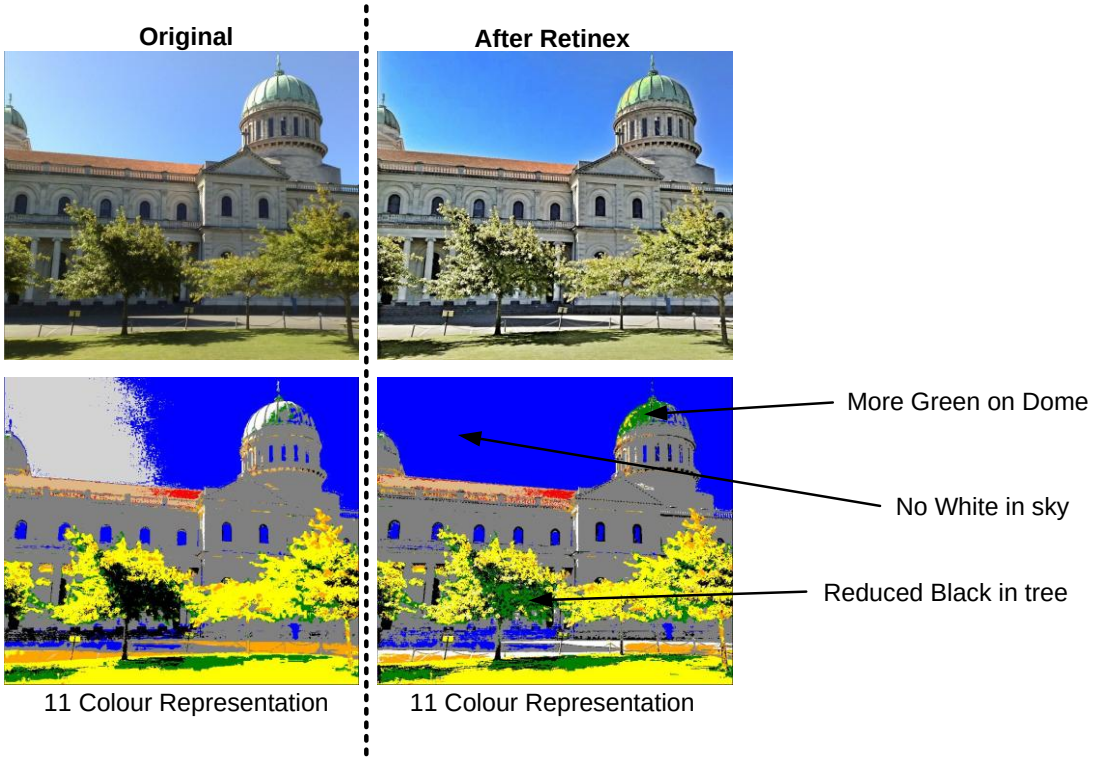


Figure 8.13: Using a Retinex Filter to Reduce Illumination Variations

8.5 Colour terms

The mapping of pixel values to colour terms is rather complex, as the perception of colour is related to many factors (Lammens and Shapiro 1993). There are issues of chromatic induction (Jameson and Hurvich 1989), which is when similar hues are judged to be different due to the contrast with surrounding colours, and effects of lighting where similar colours appear very differently depending on shadows cast onto the surface. Furthermore colour terms describe regions in colour space which are only vaguely defined, and vary depending upon the viewer.

There are many colour terms which are highly specialized and not in general use by the public (e.g. chartreuse). It was therefore necessary to first determine a list of the most popular terms in use and rate each building against these terms, generating a fuzzy colour classification (Benavente *et al.* 2006). This approach means that a target could be identified with varying strengths in a number of different colour groups, offering a degree of flexibility in any future system using the dataset, such that one user may refer to a building as "orange" while another calls it "red".

Previous research highlights 11 main perceptual colour foci (Berlin and Kay 1969, Sturges and Whitfield 1995), which are black, white, red, green, yellow, blue, brown, purple, pink, orange, and grey. Boynton and Olson (1990) found that 424 subjects could repeatedly

consistently identify the colours in this group without any confusion, therefore these 11 foci were adopted for this research.

Before assigning colour terms to an image, each colour requires a definition in colour space. The RGB values defined for each term were taken from the HP colour thesaurus, an online databank of defined colour centres constructed from people around the world (Moroney 2010), for example the most widely used red definition uses RGB values (216, 35, 44). The selected building pixels were then compared to the colour terms by translating the values into HSL colour space, before calculating their 2D Euclidean distance using the Hue and Saturation values. The process was repeated for all 11 colours giving a fuzzy classification for the building sample against the 11 colour terms. This was repeated for each sample for each target building, and the classifications were summed to produce a single building fuzzy colour set.

8.5.1 Colour Entropy

Colour entropy is a measure of the ambiguity of the assigned colour (Chuang *et al.* 2008), which may be measured by counting the number of fuzzy classes with significant values (Figure 8.14). In cases where a building has a single strong classification the colour entropy is low, and the generated natural description may include a single colour term. However where colour entropy is high a number of colour terms may be required, such as "the red-ish orange building". This fuzzy classification accommodates variation in user-formed descriptions, such that inclusion of "red" or "orange" would give the same search results. Where many buildings in view match a given selection criteria other classifications may be required to narrow the results, such as building size, or roof colour.

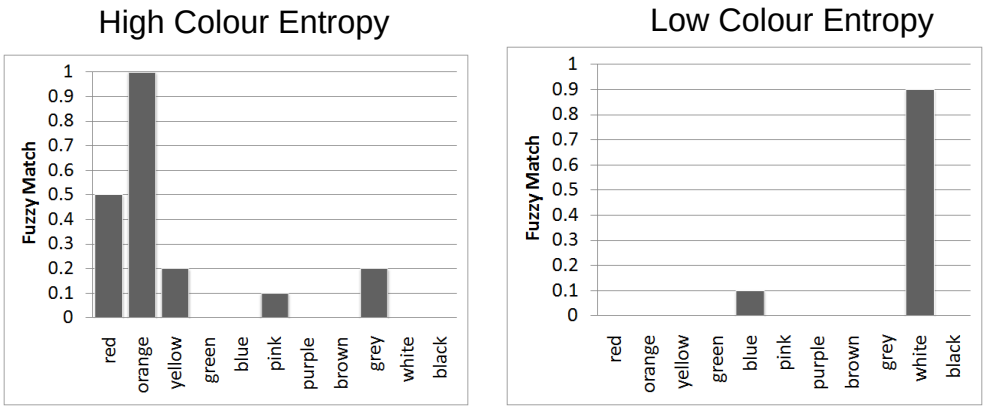


Figure 8.14: High and Low Colour Entropy Examples

8.6 Trials and Results

The system was trialled in a number of streets within Christchurch (NZ) during the summertime. The central city is grid based with a large suburban expanse, much vegetation and many gardens, providing a suitable test environment for identification and filtering of house façades from behind foreground objects such as parked cars, trees, bushes and people.

Figure 8.15 shows a selection of the identified façades, with corresponding fuzzy colour classifications. Notice that despite fairly limited views house colour could still be recovered from the views across driveways (A), and also where foreground vegetation was present (B, C). Roofs were not included in selections as a result of being non-vertical planar surfaces, and could more easily be captured from aerial imagery. Garage doors proved to be an issue however (C, D) and were generally ignored for the classification as stereo depth mapping failed to produce stable results on the very similar textured surfaces. Despite efforts to reduce selections on window surfaces there were occasions where pixels were selected (E). However on most occasions the non-window façade pixels dominated the selection, rendering the window pixels less significant in the colour classification.

To evaluate the system's performance a comparison was carried out with a sample collected from a walk along a number of streets noting house colour information. It was found that the automated procedure was able to correctly identify the most prominent colour of 33 out of 43 houses (77%) when compared to these manual values. The occasions it failed were mostly a result of incorrect colour term classification due to shadows, rather than incorrect pixel selection.

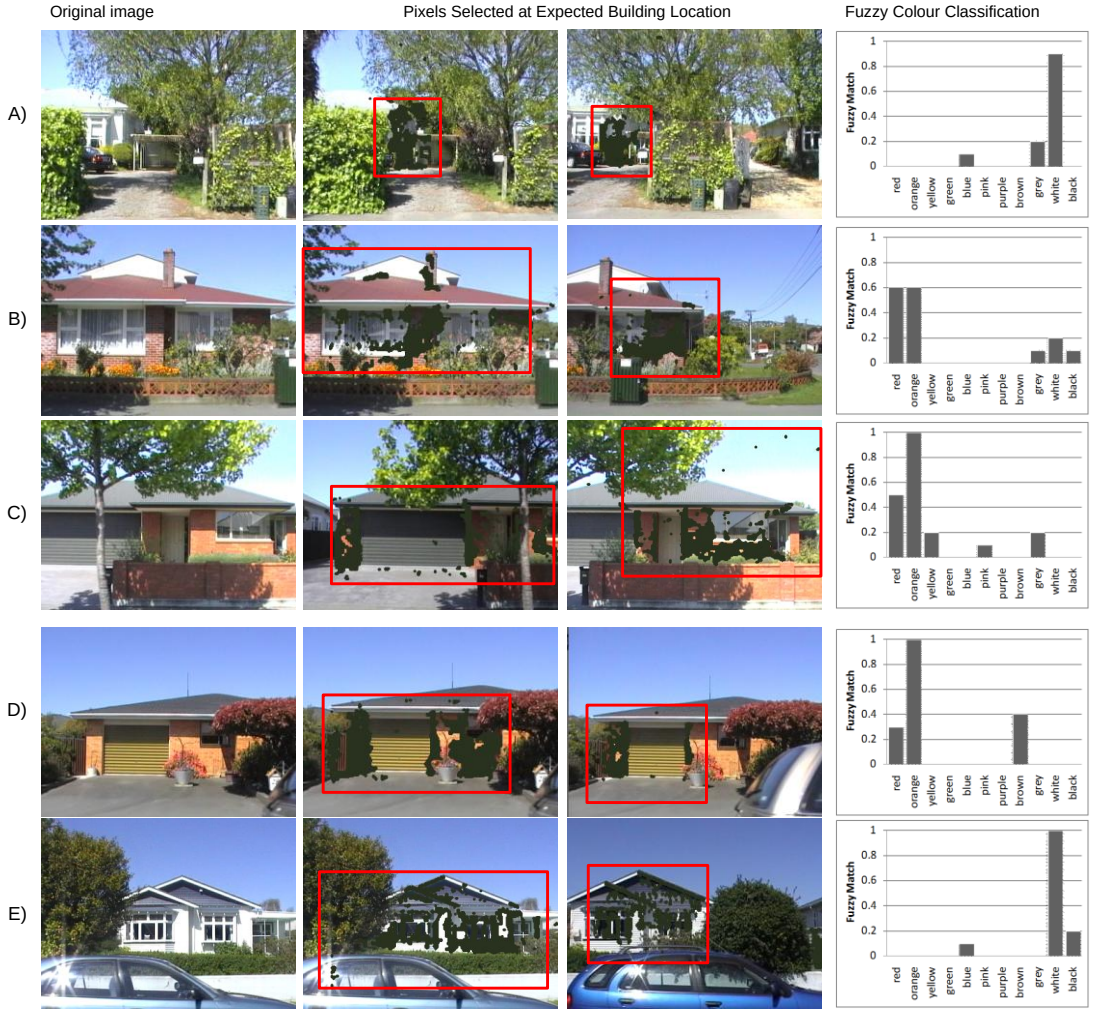


Figure 8.15: Example Façade Determination and Colour Classification. The Minimum Bounding Box for Building Pixels Selected Shown in Red.

In addition boundary wall information was gathered by selecting linear planar features. Boundary wall locations were identified correctly in 36 of the 43 targets (84%). Material recovery proved to be a more difficult task, and during trials vegetation shadows sometimes caused incorrect classifications, particularly when cast onto plain painted fences. Slat fence detection was the most reliable classification with no false positives, and few false negatives. A summary of the target houses processed are shown in Figure 8.16, including the graph of depth used to first identify the presence and location of vertical linear features in front of the house, the main planar regions, and the results of the Harris corner detection and Canny filter process.

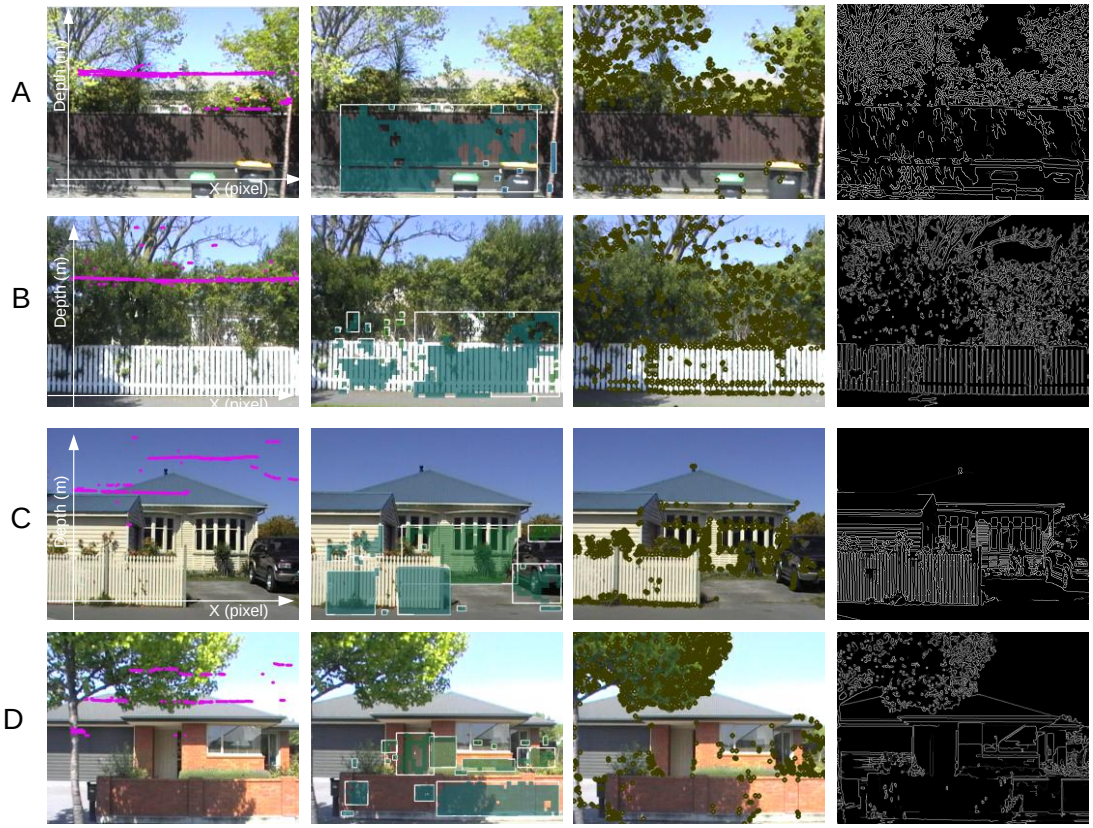


Figure 8.16: Detecting Candidate Boundary Walls and their Material

8.7 Conclusion and Future Work

When people form descriptions of buildings they often include references to attributes which are lacking in existing spatial datasets, such as façade colour and material type. In recent times there has been a dramatic rise in efforts to capture this texture and colour information from street level, however the information is locked up in images restricting its usefulness beyond visualization. To be able to extract building colour terms for use in forming natural language descriptions and for verbalization requires an ability to identify which pixels correspond to buildings by filtering out foreground objects. These pixels may then be summarized to generate the most appropriate colour term for use in forming a building description.

The work presented here demonstrates how this may be achieved using a low-tech hardware solution, available to most communities, in conjunction with spatial analysis. The processing pipeline begins by determining the expected building distance from the camera to the building by analyzing the GIS dataset. Computer vision techniques are used to build a disparity map from images captured from a moving camera, and converted to a depth map using GPS speed

information to determine the baseline distance between frames. From these inputs it is possible to filter the image for pixels at the expected building distance and generate the colour summary. Although a full evaluation of colour perception was beyond the scope of this work, the method of using a single moving camera and GPS unit to build stereo views was able to generate meaningful colour descriptions, and also to identify boundary walls and fences in images. There were however a number of issues which require attention in future versions of the system.

The use of a single camera limits the system's operation to straight road sections, as when navigating corners the rotation of the camera means parallel image views are not available. This was only a minor problem for our tests in a grid based city, but could be overcome by using the Structure from Motion approach to recover an estimate of the camera pose and orientation from which pixel depth information may then be calculated. Alternatively a multi baseline stereo camera could be used (Kang and Szeliski 1997), consisting of a set of cameras with fixed baselines, offering the advantage of an increase in operational depth recovery compared to a single stereo camera, and an ability to validate correspondences from a number of image pairs improving depth recovery robustness.

Colour term retrieval was influenced by the presence of strong shadows in images, and although the Retinex filter stage of the processing pipeline improved the situation it became evident that more sophisticated approaches would be beneficial in some cases. New techniques such as entropy minimization (Finlayson *et al.* 2009) may be worth considering for future versions of the system.

The ability to retrieve material types for both boundary walls and house facades would be beneficial when generating natural descriptions. However this proved to be difficult, for example although bricks are uniform shapes they occur in a range of colours. In theory shape detection methods would be suitable to identify them, but unfortunately Canny edge detection proved unreliable at the operating distances required. Higher quality cameras and lenses would improve this, but shadows are often cast on to the solid surfaces and may still impede these edge detection methods. Slat fences were one of the most easily recognized boundary types, as the edges were clearly defined even under strong shadow.

Currently only a single colour definition is collected per building façade or fence, however some buildings have different coloured side walls, or a multiple boundary types (eg low wall with a fence above). By exploiting the planar surface details it would be possible to divide target buildings into sections based on the direction the wall faces, and map these colour values to the GIS layer, thereby creating colour summaries for subsections of the target building. In addition the vertical boundary regions could be subdivided into those sections which appear most wall-like, and most vegetation-like to derive more complete descriptions.

Garage doors are often different colours to the main house, and would be useful additions to the descriptors. Currently they tended to be regions ignored from the summary, due to

difficulties in retrieving stable depth value for such similar textured regions. In addition window frames and doors may be segmented in the image for separate colour analysis, giving rise to very detailed house description possibilities. It is possible that windows may be identifiable as regions which change appearance with direction across views, therefore tracking a lack of consistency could be used to detect their locations.

While this study has focused on suburban regions there would be value in extending the trials to different cityscapes, including industrial and commercial regions. Colour recovery in wet conditions should also be tested to determine the effect that more reflective surfaces have on the depth and colour capture process.

The automation of building colour capture lends itself to a possible future whereby centralized colour databases are maintained from systems installed on public transport (e.g. buses, taxis), or volunteered domestic cars. In so doing colour information could be updated regularly and included in a wide range of disciplines from urban cartographic maps, to emergency response, to speech based LBS applications.

Chapter 9: Referring Expressions in Urban Environments: Generating and Comprehending Expressions using Bertin's Visual Variables

9.1 Summary

A referring expression is the part of language concerned with identifying an object through constructing a suitable descriptive phrase, for example when drawing someone's attention to a particular building in an urban scene. Typically descriptions for such features are formed from a number of attributes including colour, building material and relative position. This research demonstrates how referring expressions may be constructed for use in future speech based interfaces used in urban Location Based Services. The challenges include modelling which features are in view from any given location and how suitable descriptions may be constructed which identify the target feature, while minimising confusion with other visible features. In addition to support two directional information flow the system should be able to respond to a user's description, by determining the most likely candidate. Here a fuzzy classification system is used for each of the variables, such that the closest matches may be retrieved. The descriptive attributes used in forming the referring expressions are restricted to Bertin's visual variables, and optimal combinations are sought to ensure succinctness according to Grice's maxim of quantity.

9.2 Introduction

A conversation between two people attempting to single out a feature in view that included the phrase "Can you see that large red brick house in front of the lake?" would not seem out of place, yet the generation of such descriptions is out of reach for current Location Based Services (LBS). To generate such phrases requires a detailed knowledge of the surrounding features, a model which can determine items in view and their projective spatial relationships from the current location, and a method to find the most suitable visual attributes which may be used to form the narrative. This challenge has been referred to as the 'spatial Turing test' (Winter and Wu 2008), whereby the goal for an application's narratives is that they are indistinguishable from those made by a human.

LBSs allow users equipped with suitable mobile devices to exploit location and context when retrieving digital information on, or guidance to, items around them. As these applications become more sophisticated allowing more natural interactions, in particular as speech interfaces develop, we can expect a closer blending of digital and real worlds within our daily lives.

Speech based interfaces are usually confined to navigational devices where the task is constrained and well defined, to transfer route knowledge to the user (Hirtle and Jonides 1985, Werner *et al.* 1997), requiring only a small set of language terms. However as speech recognition tools improve their use in a range of LBS roles will increase (Michelis *et al.* 2008), particularly where the user wishes to operate in a hands-free and eyes-free way without having to re-focus attention (Francioni *et al.* 2002), such as used in the *EARS* virtual city guide (Bartie and Mackaness 2006). Aside from the mechanics of speech generation and recognition, challenges exist in how natural language phrases are generated, parsed and understood. One such challenge is in how referring expressions may be formed to guide the user's attention to a particular item within the current field of view, another is how the LBS can select the closest corresponding feature based on a user generated description. In both cases the task is to filter the features in view, after which further information can be communicated.

This chapter presents a method of evaluating the most appropriate referring expression for a feature to facilitate knowledge transfer between digital and real worlds through a speech interface, whereby digital information can be selected using real world descriptions, leading to more natural human-computer interactions. The feature descriptions are restricted to include only visual variables, meaning that no prior knowledge of the region is required (e.g. building use, age). In addition searches are limited to 'vista space' (Montello 1993), or Zubin 'C' space (Zubin 1989), that is the space around the user which can be viewed with just head movements. This does not rule out giving descriptions in advance however, such as "When you get to the junction you will see a yellow building on your left.", but does ensure that any descriptions only include reference to feature parts (e.g. a dome) which are visible.

The descriptions are generated by combining attributes from a defined list of visual characteristics, based on Bertin's visual variables (Bertin 1983). Although intended as a palette of modifications to consider when representing information in cartographic design tasks, the variables form a useful basis when describing the visual appearance of a real world object.

The chapter continues with a brief introduction to user interfaces and the importance of landmark saliency determination, before examining the role that Bertin's variables play in the context of generating referring expressions, including a discussion of suitable data sources. A method is then presented which uses fuzzy class memberships to find the most suitable description for a feature using the minimum number of attributes according to Grice's maxim of quantity (Grice 1975), ensuring succinct descriptions and minimising confusability with other items in view. Finally the task of comprehending a user's description of a feature and identifying the most likely candidates is discussed.

9.3 Background

Graphical user interfaces provide the primary interaction model in most computer systems, yet for mobile users they introduce an operational issue, demanding that the user diverts focus from the surroundings to the device's screen (Malaka and Zipf 2000). Natural Language Interfaces using speech recognition potentially provide a better alternative, enabling user support without distraction. These build on the research areas of Natural Language Processing (NLP) concerned with the comprehension of text (Manning *et al.* 1999, Allen 2003), and Natural Language Generation (NLG) which focuses on phrase formation (Dale *et al.* 1990, Reiter and Dale 1997). One of the sub-research topics in this field is Referring Expressions, which is the part of language used to describe a feature in a scene (Dale and Reiter 1995). It has particular relevancy for spatial applications as many phrases contain implicit spatial references (Kray and Porzel 2000).

Landmarks are one aspect of the environment frequently referenced, as they assist in forming mental representations of space (Hirtle and Heidorn 1993, Tversky 1993), and in way-finding tasks (Werner *et al.* 1997, Lovelace *et al.* 1999, Caduff and Timpf 2008, Winter *et al.* 2008, Duckham *et al.* 2010). Landmarks are defined as identifiable features in an environment, whose saliency may be calculated by comparing scores for particular attributes (e.g. their size) and identifying those which deviate from the mean (Raubal and Winter 2002, Elias 2003a, Elias and Brenner 2004). These are the buildings unlikely to be confused with others, either which appear very different to their surroundings (e.g. churches) or are well known major international brands (e.g. Starbucks).

While landmark saliency research focuses on finding the most recognisable features in a region, a new model is required to determine which distinguishing characteristics allow a feature to be most readily identified from the others in 'vista space' (Montello 1993). In other words, the descriptive phrase (i.e. referring expression) someone would use when attempting to identify a building to a listener. For this purpose visible attributes which are obvious without prior knowledge of a region form the most logical choice, and can be based on Bertin's (1983) set of seven visual variables which are position, orientation, size, hue, value, texture, and form. These are usually considered when displaying graphical information, but have also been used as the basis to calculate the visual component of landmark saliency (Raubal and Winter 2002, Nothegger *et al.* 2004).

For this research *position* is considered to be the location of the Feature of Interest (FOI) relative to either the observer or another visible structure (e.g. road, tree, landmark), *orientation* to be the direction the building faces, *size* to relate to the visible building façade area, *hue* and *value* are treated as wall and roof colours, *texture* is the building material (e.g. brick, wood), and

form is the overall building type (e.g house, church) including any interesting notable features (e.g. spire, dome). These are discussed in greater detail from Section 9.4 onwards.

Figure 9.1 illustrates how NLP may be used to tag the parts of speech in the sentence "the large red brick house in front of the lake facing you", in this case using the OpenNLP toolkit (OpenNLP 2011). The sentence is divided into a noun phrase (NP) which describes the target object, and a prepositional phrase (PP) which contains within it a number of NPs and PPs, giving a position with reference to the 'lake' (noun - NN). In addition the sentence has been manually parsed to show the correlation to Bertin's visual variables.

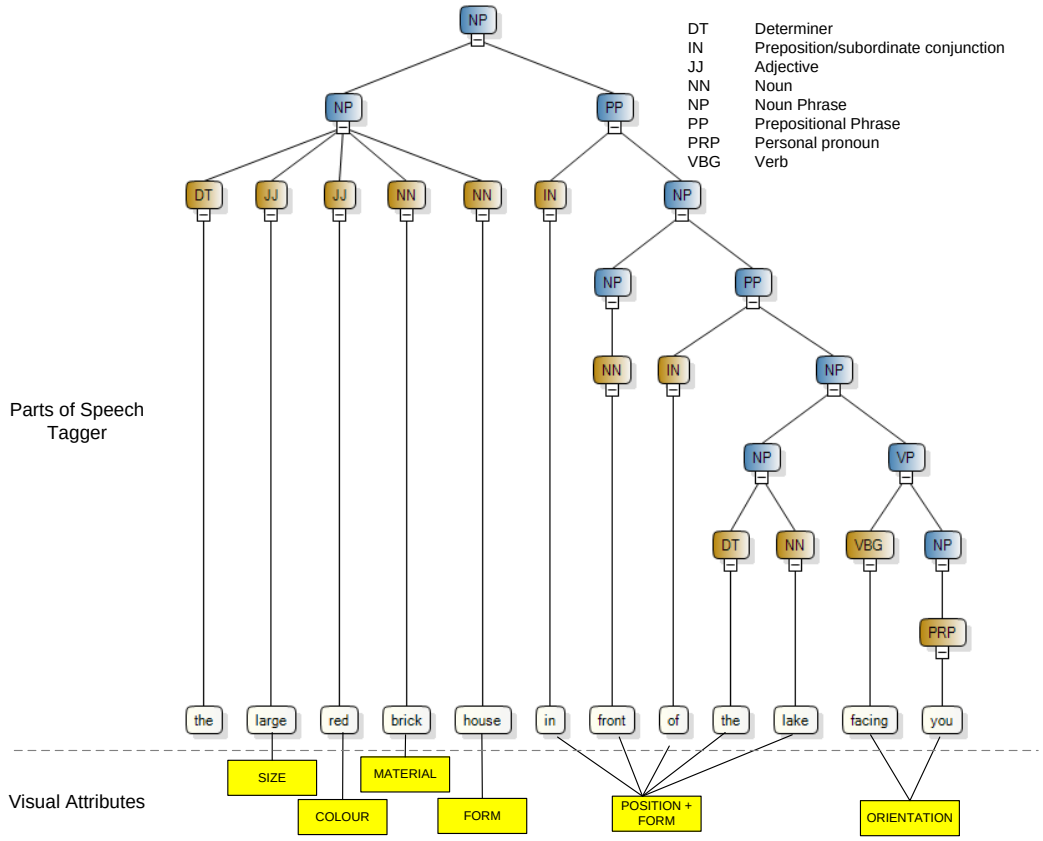


Figure 9.1: Parts-of-Speech tagging using OpenNLP, and Visual Attributes

This and further examples of how Bertin's variables are used in forming referring expressions are shown in Figure 9.2. Example A is particularly explicit using the entire set of visual variables to identify the target object, including reference to a secondary item (i.e. the lake). Example B uses only the single building *form* variable, as that description alone is enough to identify the target feature. According to Grice's maxim of quantity (Grice 1975), speech acts should be informative without providing more information than necessary. Therefore at least two bridges must have been in view for example D, making it necessary to identify the target using the *size* attribute. However it should also be noted that uttered words cannot be undone, and therefore sometimes human generated descriptions may contain superfluous descriptors which are

conveyed as the user searches for suitable adjectives (Dale and Reiter 1995). Furthermore according to Relevance Theory (Sperber and Wilson 1995) the context of the conversation implicitly provides a framework which can remove the need to generate full descriptions in every sentence. Therefore during a request for information, for example clarification on which cafe to meet a friend, the system need only reply with a minimal description which assumes both parties have pre-conceived contextual knowledge (Example G). Even without considering previous dialogue an LBS would be designed to carry out a certain role, such as a virtual city guide, and have an imposed explicit framework.

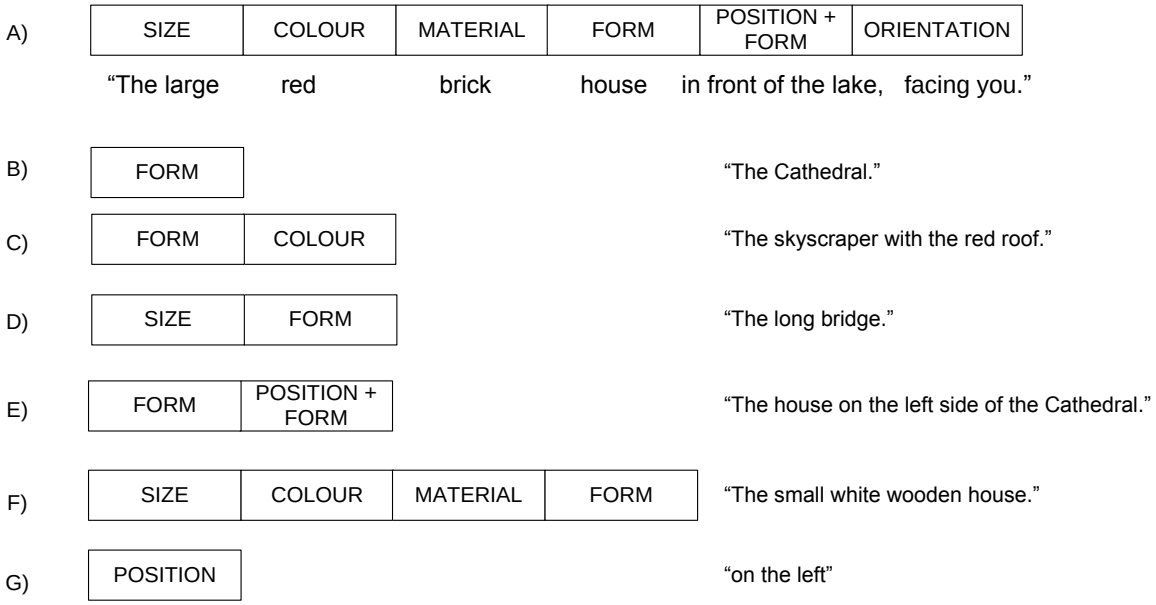


Figure 9.2: Bertin's Variables as Description Building Blocks

Related previous research has demonstrated the use of NLG in navigational map tasks (Varges 2005), and the importance of spatial relations in referring expressions (Viethen and Dale 2008). Spatial datasets have also been used in the automatic generation of appropriate image captions by comparing a photograph's location and orientation against known land cover, landmarks, road and other datasets (Tomko *et al.* 2009). Such research has focussed on rural regions, but demonstrates the capability of using spatial resources to construct suitable descriptive texts. The demands for generating FOI descriptions in urban regions are similar but require higher fidelity data sources.

The model developed here is in anticipation of such data sources providing a complete resource for each of Bertin's variables across the entire urban region. Developments towards this include complete textured 3D city models (Brenner *et al.* 2001, Frueh *et al.* 2004, Wang and Li 2008) and new techniques to capture building textures while removing the unwanted foreground

objects (e.g. cars and trees) (Forsyth *et al.* 2008), or fill in the background using images from alternative angles (Yi-Leh *et al.* 2010). A sensor fusion approach has also been successful in collecting urban form where laser ranging equipment is used in conjunction with cameras to collect textures (El-Hakim *et al.* 1998, Pylvänäinen *et al.* 2010) as demonstrated on websites Yell (Yell 2010) and C3 (C3 Technologies AB 2009). Furthermore FOI c*olour* terms may be collected from street level using simpler stereo vision techniques combined with building footprint maps (Bartie *et al.* 2011), and *position, orientation, size* information can be sourced from LiDAR data combined with building polygon data. Feature *form* may be sourced in part from LiDAR surface model interpretation (Brenner *et al.* 2001, Rottensteiner and Briese 2002), or from user contributed data sources (Espinoza *et al.* 2001, Goodchild 2007). An example of this wisdom of the crowds approach is demonstrated in identifying the most commonly photographed parts of cities, and determining the interesting parts of buildings (e.g. fountains, statues, spires, domes, gargoyles, clock towers, entrance gates) from public image sharing websites (Simon and Seitz 2008). Through combining these initiatives rich urban datasets are becoming available which will be suitable for performing sophisticated referring expression generation.

9.4 Fuzzy Classification for Bertin's Variables

To determine the most appropriate combination of visual variables a scoring system is used that evaluates the strength of a referring expression. This is done by dividing each variable into a number of sub-classes and assigning each feature a corresponding fuzzy membership value. For example *size* may be defined according to a "largeness" scale, to which a FOI may be assigned a value of 0 to indicate it is very small, 1 to indicate it is very large. The feature which is furthest from its nearest neighbour along this scale is considered to be the most identifiable, this is not the furthest from the mean, nor an outlier at either extreme, but instead a point which is distinguishable from other clusters. Figure 9.3a illustrates this for a single visual characteristic whereby five features are in view, and values for a sub-class plotted along an axis. Items 1 and 3 may be expressed using the superlatives smallest and largest respectively, while item 2 is the furthest from its nearest neighbours. Considering the *size* sub-class example then item 1 may correspond to a shed, item 2 a house, and 3 a skyscraper. Although each set is identifiable the difference between the size of the two sheds, or the size of the 2 skyscrapers may be comparatively small and difficult to distinguish. However the house set is distinct, making item 2 the most identifiable single feature in view.

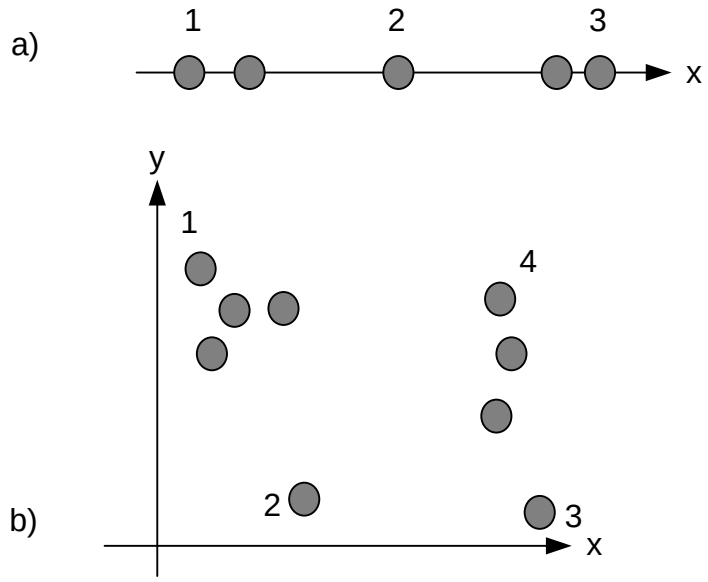


Figure 9.3: Attribute values in n-dimensional space – item 2 has the furthest Nearest Neighbour

Combinations of values are also considered as shown in Figure 9.2 (e.g. form and colour), by determining neighbours in *n*-dimensional space, until the strongest referring expression combination is found. Figure 9.3b illustrates this for a 2-dimensional space, whereby item 1 has the highest y-axis value, item 3 the largest x-axis value, item 4 the largest combination of both x-axis and y-axis values, and item 2 is the furthest from its nearest neighbour. This approach can be extended into *n*-dimensions such that the distance to a nearest neighbour in *n*-dimensional space may be found.

The list of subset classes used for each visual variable will depend on the data availability and may be adapted to suit specific LBS roles. The rest of this section explains in greater detail how the subset classes are formed and used to form descriptions, before demonstrating their use in a worked example in Section 9.5.

9.4.1 Positional information

Positional information may be defined in many ways by describing the relationships between the FOI and its surroundings, or relative to an observer. Relationships are rarely described in metric space (e.g. 123.7m at 54 degrees) but instead usually refer to with topological (Egenhofer and Herring 1990, Egenhofer 1991, Clementini and Di Felice 1995, Egenhofer and Franzosa 1995) or projective spatial relationships (Clementini and Billen 2006). For example a paddling pool may be described as being "inside the park" using a topological relationship, or "in front of the swings" using a projective relationship. Alternatively a house may be described as "on your left" by referencing the observer.

Referring expressions which use topological operators such as intersect, within, and contains (Egenhofer and Herring 1990, Riedemann 2004), require access to a number of underlying spatial layers (e.g. parks, city boundaries). From this it is possible to generate a list of Boolean or graded topological memberships for each FOI, for example a building on campus would be allocated a score of 1 for the "within" topological operator when compared to the campus boundary layer. The best referring expressions formed solely from topological expressions are those which result in the minimum number of confusable targets, evaluated as the operator set with the lowest occupancy. Typically topological operators will be used in conjunction with other visual variables to form unique descriptions, such as "the white house on the hill". The topological relationships between buildings and the underlying layers are static, meaning they could be pre-calculated and stored with each FOI.

Projective relationships, however, are dynamic as they include the observer's viewpoint (Hernández 1991), and may be calculated using a visual exposure model to return a collection of the visible parts of FOIs (Llobera 2007b, Bartie *et al.* 2010). The relationships between these visible parts are then classified either according to a division of space around the user, or by referencing a landmark, using the 5-intersection model (Clementini and Billen 2006). This defines 5 regions which are referred to as before, after, left side, right side, and between. A graded approach may be adopted and relationships may be used in combination. For example, "the house is before and left of the bridge" may receive scores of 1 for the projective relationship *before*, and 0.5 for *left of*, indicating the house is entirely in front of the bridge but not entirely left of it with some overlapping parts (Bartie *et al.* forthcoming).

For this research features are first defined using topology, then for each subset (e.g. those in the park) with respect to the user (e.g. on your left), then finally with respect to the most prominent feature (i.e. landmark with the highest saliency). In each case the ambition is to find a distinctive definition, but if none is possible then orientation and other attribute combinations are considered.

9.4.2 Orientation

Typically the building's orientation is calculated by establishing the longest axis using the rotating callipers method (Shamos 1977) or equivalent technique, yet this does not reflect the building's observable orientation from street level. Instead a method is required which can establish if a wall is approximately perpendicular to the observer (i.e. facing observer) or roughly parallel to the viewed direction. Walls facing the observer are generally more easily noticed as they occupy a larger field of view than oblique walls (indicated as A in Figure 9.4) and therefore are more suitable additions to a referring expression.

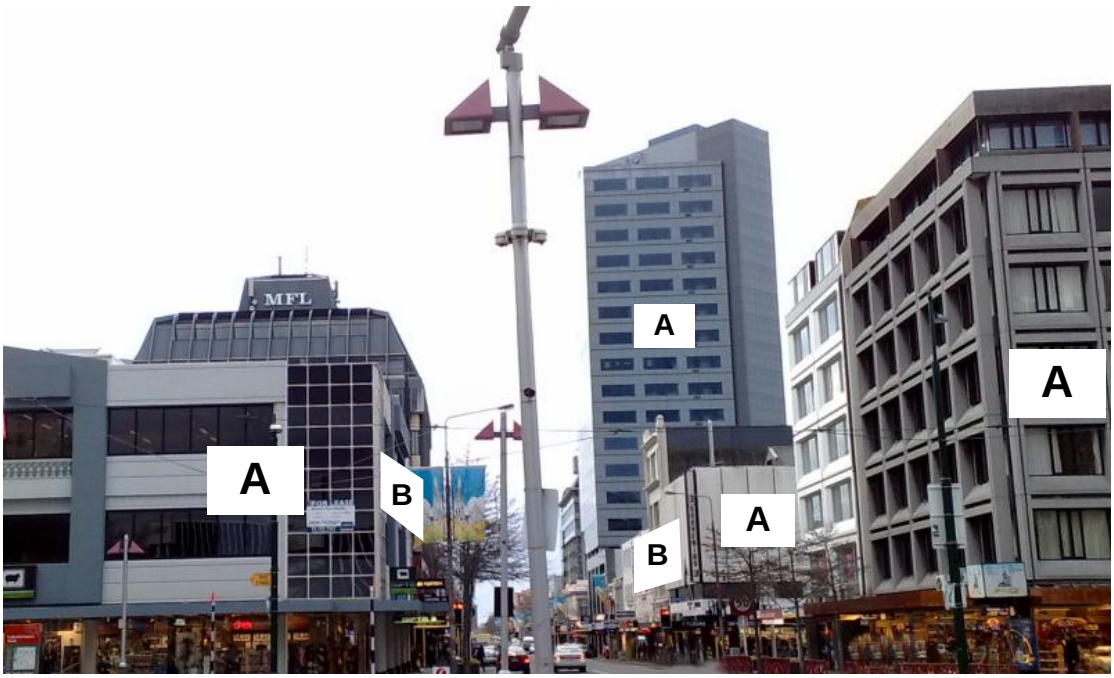


Figure 9.4: Orientation of Exterior Walls: Main component approximately:-
(a) Perpendicular to observer (b) Parallel to Observer's Viewing Direction

Simply testing the angle of incidence of each ray with the defined polygon boundary fails to capture any prominent faces within the region, and therefore it is necessary to consider the interaction between the ray and the aspect of each target feature cell. Aspect values may be generated from a Digital Surface Model (DSM), classifying each cell as flat or one of 4 direction categories (i.e. N/S, NE/SW, E/W, SE/NW). The visibility results from this procedure are displayed in Figure 9.5 for the viewpoint shown in Figure 9.4. The cells visible (Figure 9.5b), which consist of building edges and the taller sections visible above closer lower buildings, are then applied as a mask to the aspect values. By comparing the incoming lines of sight cast from the observer to these visible cells it is possible to summarise the average relative ray incidence angle per building. These are expressed as the number of visible cells with a ray incidence greater than 45 degrees over the total number of cells visible for the FOI, giving a perpendicularity rating between 0 and 1 for the feature overall. Figure 9.5c shows the summaries for this example, whereby FOIs C,D,E are mainly parallel to the observer, while A,B,F and G have strong perpendicular components. A majority of 45-90 degree interactions (e.g. G – the tall skyscraper on the right of Figure 9.4) indicates that the most noticeable part of the FOI is perpendicular to the observer's viewing direction and forms a useful orientation descriptor.

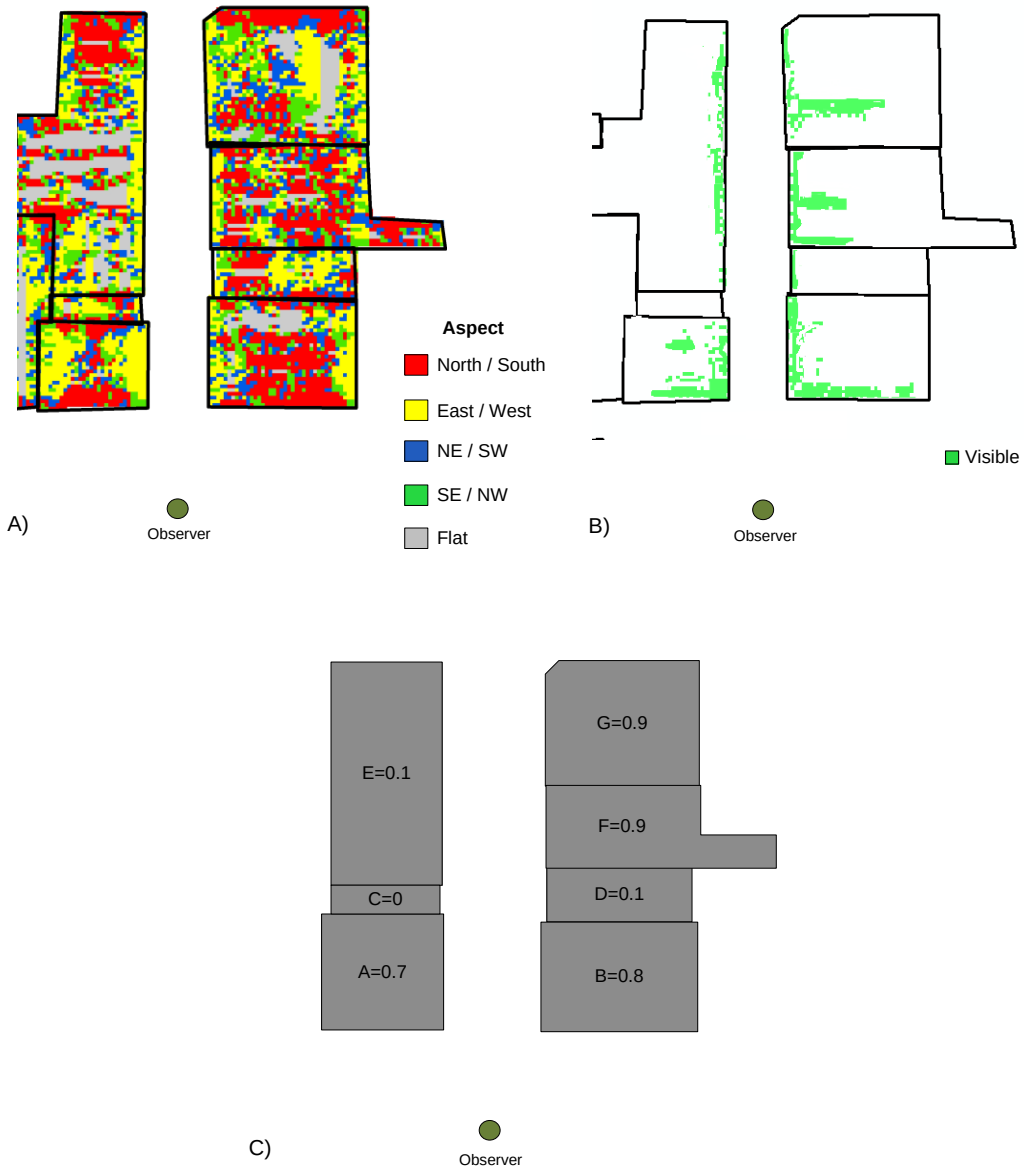


Figure 9.5: (a) Aspect classes for Calculating Wall Directions (b) Visible FOI Parts (c) Visible Perpendicular Wall Class Membership per FOI (where 1.0 indicates presence of wall perpendicular to view)

9.4.3 Size

Building size is a reflection of the floor area and building height, with consideration given to the current viewpoint. Two outputs from the visual exposure model which describes this are the *façade area,* calculated by establishing the total of vertical extents visible for DSM cells within the designated FOI boundary, and the *perceived area* which also considers the viewing distance and angle (Bartie *et al.* 2010). Figure 9.6 illustrates the differences between these metrics, whereby the *façade area* for some distant buildings indicates they are visually impressive (e.g.

skyscrapers) something the observer will be aware of despite their distance, while the *perceived area* reflects that closer buildings appear larger to the observer and dominate the field of view.

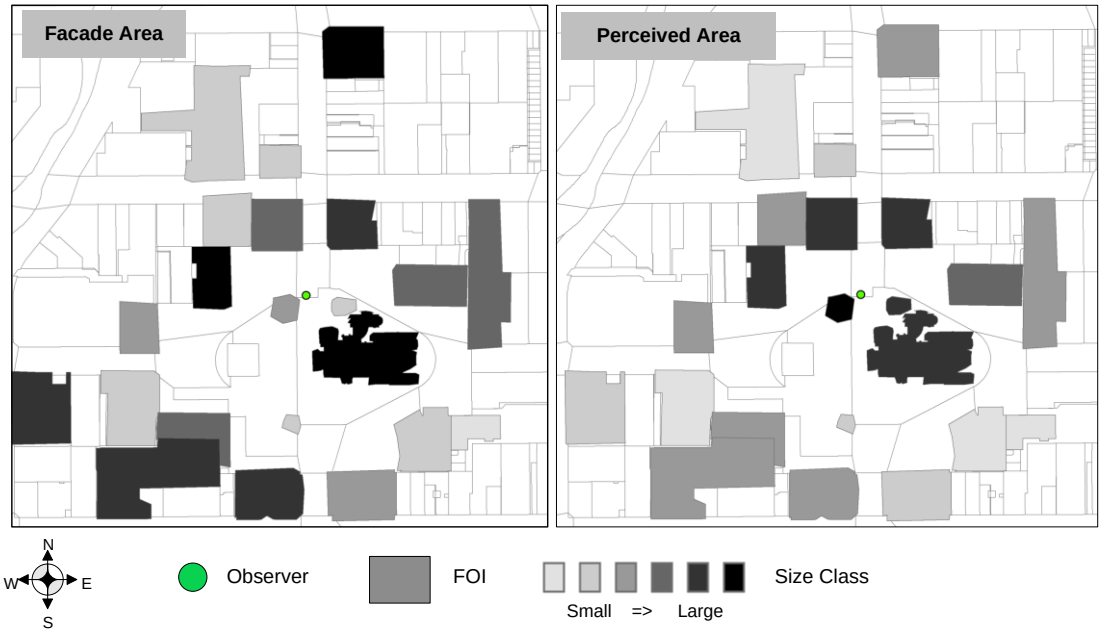


Figure 9.6: FOI Size - façade area and perceived area

Expressing all the visible FOIs by either façade or perceived area as a ratio of the area of largest visible item allows for an appropriate membership value to be assigned against a 'largeness' score. Building size may also be expressed using axial based superlatives such as the tallest/shortest, deepest/flattest, or narrowest/widest. However the use of these terms are scale dependent, as a skyscraper 5 metres taller than others in the region would not be noticeable to a ground based observer, therefore minimum tolerances are required before such superlatives can be used, something which should be established with user trials and is not considered further at this time.

9.4.4 Colour

Colour is a valuable descriptor which can be used to rapidly filter possible candidates in a scene. It has been found to be useful in object recognition (Wurm *et al.* 1993), discriminating features (Ennesser and Medioni 1995), way-finding indoors (Brown *et al.* 1997), and outdoors (Helvaciog lu and Olguntürk, Amri Musliman *et al.* 2010). There are many colour terms which are highly specialised and not in general use by the public (e.g. chartreuse), so it is necessary to first determine a list of the most popular terms in use and rate each building against these to give a fuzzy colour classification (Benavente *et al.* 2006). Previous research highlights 11 main

perceptual colour foci (Berlin and Kay 1969, Sturges and Whitfield 1995), which are black, white, red, green, yellow, blue, brown, purple, pink, orange, and grey.

Figure 9.7 shows an example of how colour collection may be automated, whereby stereo images were used to calculate the distance from the camera for each pixel. These distances were compared to that measured from the camera location to a map of building footprints, from which the pixels at the corresponding distance were selected and classified according to the 11 colour foci. In these two examples the selected FOI pixels reveal in the first case a mostly orange/red house, and in the second a mostly white house (Bartie *et al.* 2011). Through the use of a fuzzy colour classification approach the model is able to accommodate the variety of descriptions generated by users, so that one person may refer to a house a 'red' house while another calls it 'orange'.

Additional colour information may be retrieved for window, door and boundary walls if desired, as well as roof colours from aerial imagery. As before the description may only be suitable in forming referring expressions when combined with other visual attributes whereby the number of confusable target candidates is reduced.

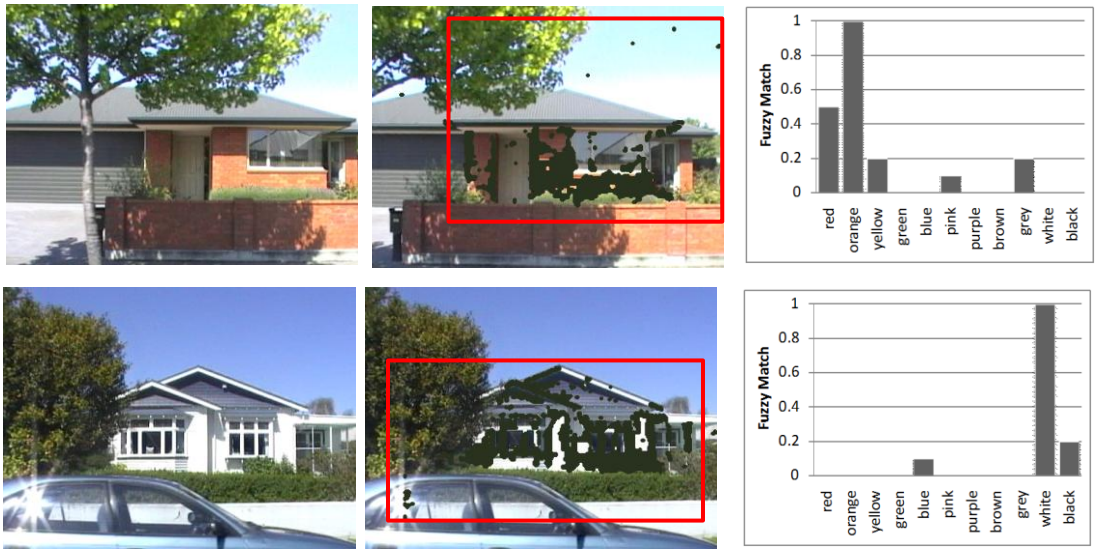


Figure 9.7: Fuzzy Colour Classification Examples – showing from left to right the original image, selected façade pixels, and final colour classification.

9.4.5 Texture

Texture may include building materials such as brick, glass, plaster, stone, wood, steel or concrete. However the material type may not be obvious depending on the finish, and is one of the most difficult attributes to collect using automatic techniques (Pajdla *et al.* 2004).

Texture information is only relevant when easily identifiable from a distance and usually in support of a colour attribute, such as the 'red brick house'. Fuzzy memberships are possible by

classifying each FOI's closeness of fit to a defined model, for example a model for bricks might be defined as a regular red/brown/orange rectangle wider than they are tall. A surface composed of these may receive a score of 0.9 brick-like to signify the strong presence of such shapes. Similarly other models may be developed to accommodate a range of textures and surfaces. Until the automation of this is further developed the most suitable approach may be to manually classify buildings using a Boolean classification.

9.4.6 Form

The construction type is referred to as FOI *form*, such as building (e.g. house, bungalow, skyscraper, church), statue, or bridge. These are fixed definitions, but still benefit from the fuzzy classification strategy, for example allowing an FOI to be classed with a strong bungalow membership as well as a weak house membership. In addition notable key features are considered under the *form* attribute, such as spires, porches, domes and turrets. Signage could also be considered as a key feature, but urban text recognition is made more problematic by the wide variety of fonts and logo designs (Slawski 2007), therefore data are limited.

Depending on the user's viewpoint different aspects of a FOI may be observed, as shown in Figure 9.8, therefore to ensure that referring expressions only include reference to visible key features it is necessary to make a number of adaptations to the visibility model as explained below.

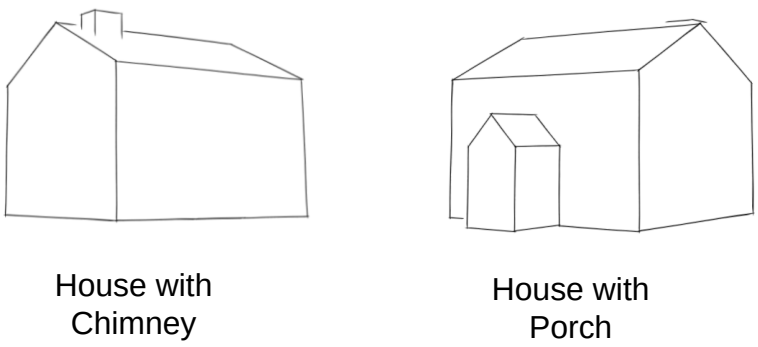
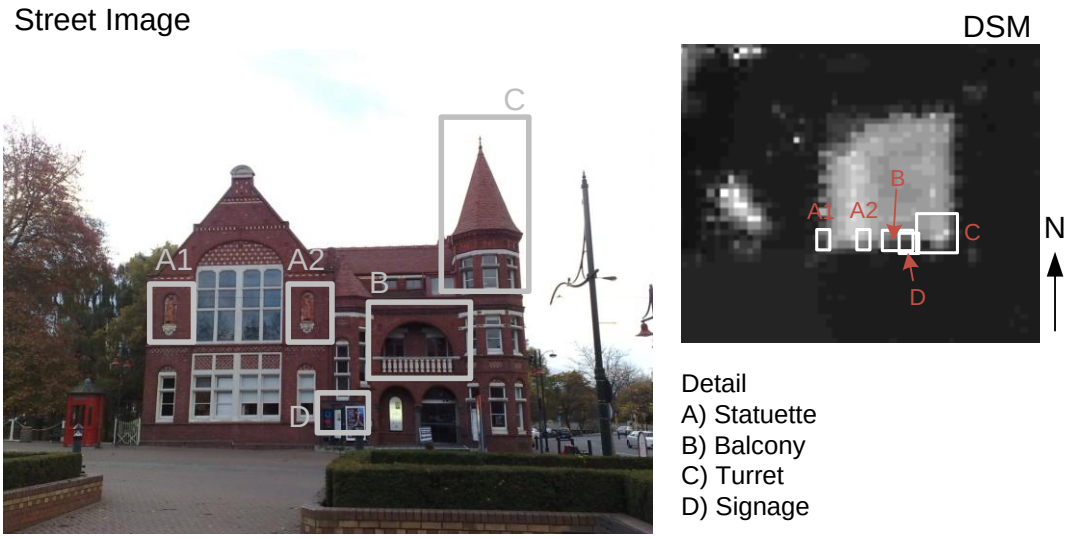


Figure 9.8: Viewing Angle and Form

Visibility modelling returns information on which cells in the DSM are visible without denoting which sides of the cell are visible. However in some instances this information is needed to determine whether a particular feature on the side of a building (e.g. a sign) can be viewed. Therefore key features are mapped to a set of DSM cells within the FOI boundary, with their viewing restrictions added as necessary depending on the key attribute type. For those on the side of buildings, such as balconies, entrances, or signage, a viewing angle restriction is added preventing positive identification when the cells are visible from the opposite side of the

structure. This ensures that referring expressions will only mention those key attributes when viewed from in front of the feature. In other cases, such as a turret or spire, the viewing angle does not need be restricted and may be mentioned whenever the corresponding cells are visible. Limitations to viewing height are also required in certain cases so that, for example, a description of an entranceway is not included if the observer may only view the top of the FOI across rooftops. Finally viewing distance restrictions may be added, which are particularly useful for limiting reference to signage viewed from far away. An example of how the supplementary viewing restrictions may be assigned to an FOI is shown in Figure 9.9, which shows features denoted on a building and the corresponding imposed viewing limitations. These notable features may be added to a referring expression to assist in its identification, such as the "building with a turret".



Detail	**Example A1,A2**	**Example B**	**Example C**	**Example D**
Attribute Type	Structure	Structure	Structure	Signage
Attribute Value	Statuette	Balcony	Turret	"Our City"
Min Elevation (m)	4	2	Any	0
Max Elevation (m)	6	5	Any	2
Facing (degrees)	270 => 90	270 => 90	Any	270 => 90
Maximum viewing distance (m)	50	50	Any	10

Figure 9.9: Viewing Restrictions for Key Attributes for a Feature of Interest

9.5 Referring Expression Generation

The purpose of description generation is to form a sentence which uses the most appropriate visual attributes enabling a feature to be identified from a given location. The attribute selection is controlled in part by data availability, as attributes may only be considered for inclusion if a value is available for all of the items in view at one time. For example "the yellow house" implies that all visible house colours are known and only a single yellow one is in view. If instead colour information is not known for all visible houses then this expression should not be generated.

In the case of description generation there are two methods required, the first is to find the most salient object in the current view (e.g. landmark), and the second to determine the most suitable description for a given object in the current scene. The first function is similar to previously presented models of scoring landmark saliency (Raubal and Winter 2002, Elias 2003a, Nothegger *et al.* 2004), while the second is a novel function which determines the strongest descriptive variable to use for each of the FOI in view.

The approach here is to classify each FOI using a fuzzy membership assigned to each attribute class. For a full implementation many subtypes would be evaluated as outlined in Table 9-1 but for the purposes of this example the classes were limited to Position relative to the observer (Left to Right), Colour (Red, Orange) and Size (Visible area).

Table 9-1: Visual Attributes and their Sub-classes

Attribute	**Sub-Class**
Position (topological)	[Within] [Crosses] [Touches]
Position (ref. observer)	[Left to Right]
Position (ref. landmark)	[Left to Right] [Above to Below] [Before to After]
Orientation	[Facing]
Size	[Façade Area] [Perceived Area]
Colour	[Black] [White] [Grey] [Red] [Orange] [Yellow] [Green] [Blue] [Brown] [Purple] [Pink]
Texture	[Brick] [Stone] [Wood] [Glass] [Steel] [Concrete] [Plaster]
Form (class)	[Bungalow] [House] [Skyscraper] [Church] [Statue]
Form (notable features)	[Turret] [Spire] [Chimney][Dome][Statuette] [Balcony] [Signage]

Categorical data are handled by dividing the class into a number of types, for which the value is compared and assigned a membership value. For example the colour attribute is divided into the 11 colour members, and visible portions of the façade are compared to these colour foci allocating a membership to each (e.g Red(0.8), Orange(0.4), Yellow(0.1), Green(0), etc),

similarly textures (e.g. brick(0.9), stone (0.1), wood(0.1), glass (0)), and form (e.g. church-like (0.8), house-like (0.1)) are handled in this way.

Numerical data are handled by setting an axis definition and ranking the visible FOIs accordingly. For *size* this may be a 'largeness' axis, to which perceived visible areas are ranked, such that the biggest FOI would receive a membership of 1, and the smallest a value of 0. Buildings of size in between would be allocated a value proportionate to their size compared to the largest building. What follows is a worked example of how a subset of these attribute classes may be used to determine the most salient object in a scene (section 9.5.1) and then the best descriptor for a nominated FOI (section 9.5.2).

9.5.1 Most Salient Object - Worked Example

The most salient object in any scene may be considered as a landmark, which can be included in a referring expression as the basis from which to navigate the user's gaze towards less salient features, as in "The brick house left of the *statue.*". This may be calculated as follows,

$$r = max\big(min(d_{ij})\big) \,/\, sec\,\big(min(\,d_{ij})\big) \qquad \textit{equation 9.1}$$

where r is the ratio of the maximum (*max*) nearest neighbour (*min*) distance in the matrix for all items (d_{ij}), and the second largest value (*sec*) in the same matrix. The most salient item is that which exhibits the greatest ratio (r), as it has the greatest separation from all other items in view.

The following simplified worked example shows how this may be used with fuzzy classes to determine the strongest description for an FOI. Here for simplicity the attribute list is limited to show only two colours from the list of 11, and attributes for size and position.

Imagine the scene shown in Figure 9.10 where an observer can view five buildings, the attributes for which are displayed in Table 9-2. To identify the most outstanding building in the scene each of the attributes is compared to the other values, determining the difference as a distance measurement. This can be considered as identifying the maximum distance to the nearest neighbour in n-dimensional space, where n defines the number of attributes used to construct the referring expression. The FOI with an attribute combination which results in the largest distance to its nearest neighbour is considered the most salient, or distinctive.

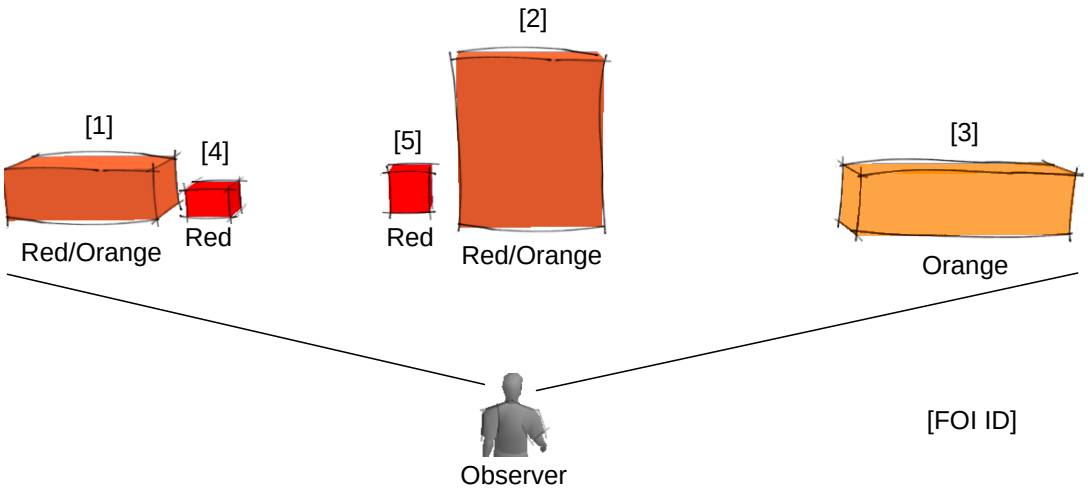


Figure 9.10: Observer's View of a Scene

Table 9-2: Most Salient Object in a Scene

	Position		COLOUR		SIZE	
ID	Left to Right	Rightmost Membership	RED Membership	ORANGE Membership	Area	Largeness Membership
1	L	0	0.8	0.4	200	0.8
2	M	0.5	0.8	0.5	250	1
3	R	1	0	1	220	0.88
4	L	0.1	1	0	100	0.4
5	M	0.4	1	0	115	0.46

A visual representation of the position and size attributes is shown in Figure 9.11, with lines indicating the distance to neighbours for FOI 2, whereby the nearest neighbour is identified as FOI 3. In other words FOI 2 and FOI 3 are the most similar in the scene comparing size and position, and therefore the most likely candidates to be confused for each other. By finding the maximum distance to the nearest neighbour for all variable combinations, so the FOI which is the hardest to be confused with any other in the scene can be identified.

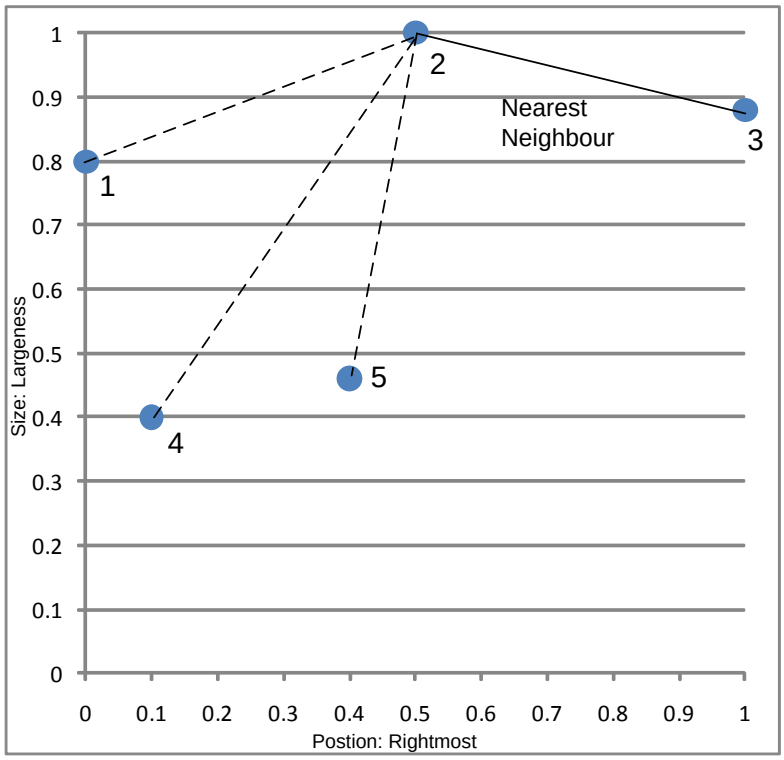


Figure 9.11: Graph of Position and Size, Indicating Distance from FOI 2 to Other FOIs

Table 9-3 shows a subset of all the combinations forming the result matrix of distances, as well as the minimum distance between items for each item in each category, in other words the nearest neighbour. For example considering just the *position* attribute FOI 3 has the greatest separation from its nearest neighbour, signifying that position is a valid choice in forming a descriptor for this FOI. Similarly FOI 2 is identifiable from the other buildings using its *size* attribute, as this is the building with the greatest separation from the next largest building.

Table 9-3: Result Matrix for Most Salient Object (subset of results shown)

Position	(ID)	1	2	3	4	5	Minimum	Max. of Min.	1st:2nd Ratio
(ID)	1		0.500	1.000	0.100	0.400	0.100		
	2	0.500		0.500	0.400	0.100	0.100		
	3	1.000	0.500		0.900	0.600	0.500	0.500	5.000
	4	0.100	0.400	0.900		0.300	0.100		
	5	0.400	0.100	0.600	0.300		0.100		

Size	(ID)	1	2	3	4	5	Minimum	Max. of Min.	1st:2nd Ratio
(ID)	1		0.200	0.080	0.400	0.340	0.080		
	2	0.200		0.120	0.600	0.540	0.120	0.120	1.500
	3	0.080	0.120		0.480	0.420	0.080		
	4	0.400	0.600	0.480		0.060	0.060		
	5	0.340	0.540	0.420	0.060		0.060		

Colour	(ID)	1	2	3	4	5	Minimum	Max. of Min.	1st:2nd Ratio
(ID)	1		0.100	1.000	0.447	0.447	0.100		
	2	0.100		0.943	0.539	0.539	0.100		
	3	1.000	0.943		1.414	1.414	0.943	0.943	9.434
	4	0.447	0.539	1.414		0.000	0.000		
	5	0.447	0.539	1.414	0.000		0.000		

Position +Size	(ID)	1	2	3	4	5	Minimum	Max. of Min.	1st:2nd Ratio
(ID)	1		0.539	1.003	0.412	0.525	0.412		
	2	0.539		0.514	0.721	0.549	0.514	0.514	1.000
	3	1.003	0.514		1.020	0.732	0.514	0.514	1.000
	4	0.412	0.721	1.020		0.306	0.306		
	5	0.525	0.549	0.732	0.306		0.306		

The ratio between the highest and second highest places from each table is used to normalise results, accounting for the variation in the number of attribute inputs, and determine the strongest attribute combination for forming referring expressions. Table 9-4 shows the results (i.e. *r* values calculated using equation 9.1), indicating the most resilient description would be formed using the *colour* attribute alone for FOI 3, as it is the only *orange* item in the scene. *Position* is the second strongest descriptor for FOI 3 as it is the only item on the user's right, while the *size and colour* attributes make the third strongest description. Notice too that *size* alone is useful in identifying FOI 2, but that separation from the second place item is relatively small, giving this

descriptor a fairly low strength. Therefore in the current scene the most salient object would be classed as FOI 3, which is identifiable foremost by its colour, but also by its position.

This matrix serves another purpose which is to reveal the most likely confusable items based on an attribute. For example considering colour for FOI 1, it can be seen that FOI 2 occupies a fairly similar colour space (0.1), while item 3 is very different (1.0). Therefore we can assume that it is unlikely for a user to confuse items 1 and 3 based on colour, but more likely items 1 and 2 may be confused.

The method for determining the most suitable descriptor for a particular FOI is discussed in the next section.

Table 9-4: Strongest Descriptor (r values)

Attribute Combination	Ratio (r) 1st:2nd	Top ID
Colour	9.434	3
Position	5.000	3
Size+Colour	4.253	3
Position + Colour	2.094	3
Position + Size + Colour	1.961	3
Size	1.500	2
Position + Size	1.000	2,3

9.5.2 Best Referring Expression for Nominated FOI – Worked Example

Generating a good description which uniquely identifies a FOI may be achieved by combing attributes until a single entry is identifiable. To ensure brevity Grice's maxim of quantity is adopted (Grice 1975), which states that the minimum of information should be conveyed to achieve this goal. This is achieved by determining the strongest normalised value using an exhaustive search of variable combinations, built from the minimum number of attributes. Normalisation is necessary to remove the effect of higher values occurring when greater numbers of attributes are used, and achieved by expressing each membership as a ratio of the maximum for the attribute combination. This may be shown as,

$$r_{n=1}^{i} = min(d_{nj}) / \, max(d_{ij}) \qquad \textit{equation 9.2}$$

where r is the ratio of the nearest neighbour for an item (min (d_{nj})) and the largest distance to a nearest neighbour (max (d_{ij})) from all possible items using the same attribute combination.

Applying this formula for the scene shown in Figure 9.10 and corresponding values from Table 9-2, an output is generated as displayed in Table 9-5. The results indicate that FOI 1 is the largest item on the left best described using the size and position attributes, while for FOI 2 is the largest item in view (*size* attribute). FOI 3 was previously identified as the most salient object in view, and can be described in many ways. Here Grice's maxim would dictate using the lowest number of attributes, however it is also possible to determine the most dominant attribute combination as before (Table 9-4) by using the ratios of highest and second placing. This results in FOI 3 being described by its *colour*, as the only orange item. Finally FOI 4 and FOI 5 are most adequately described using *position and size* attributes, being the smallest items on the left and in the middle respectively.

In summary, the model output shows FOI 1 as "large building on the left", FOI 2 as "the big building", FOI 3 as "the orange building", FOI 4 as "the small building on the left", and FOI as "the small building in the middle".

It is worth noting that although the largest object in view, FOI 2, is identified through its size attribute alone, the smallest item was not identified by size. This is a result of the confusion which may occur when considering FOI 4 (smallest at 100 sqm) and FOI 3 (110 sqm), as the percentage difference between areas is minimal. However the largest item FOI 2 (250 sqm) has a significant size increase from the second largest item (FOI 3), and is therefore considered as a strong descriptor.

Table 9-5: Determining the Most Suitable Descriptors for a FOI (P=Position, C=Colour, S=Size)

ID	P	S	C	P+S	P+C	S+C	S+C+P
1	0.2	0.666667	0.106	0.801851	0.429198	0.235128	0.50978
2	0.2	1	0.106	1	0.477567	0.235128	0.50978
3	1	0.666667	1	1	1	1	1
4	0.2	0.5	0	0.594987	0.280976	0.063092	0.284747
5	0.2	0.5	0	0.594987	0.280976	0.063092	0.284747

The next section briefly examines the procedure required in comprehending a user generated description, and determining the most likely match for the list of items in view.

9.6 Description Comprehension

The sequence of events for responding to a question from the user is outlined in Figure 9.12, beginning with the formation of a description in the user's mind, to transmission as speech, and recognition and parsing by the LBS. Here the recognition engine and ability to parse the speech are treated as mature technologies, so that discussion can focus on the target identification

aspects. It is also assumed that the user only refers to visual variables for which data are available.

The parsing process is not discussed here, but would need to identify the set of visual attributes mentioned in the narrative in a process similar to that shown in Figure 9.1. The fit between all modelled visible FOIs and the described item are then compared, resulting in the closest match as demonstrated below.

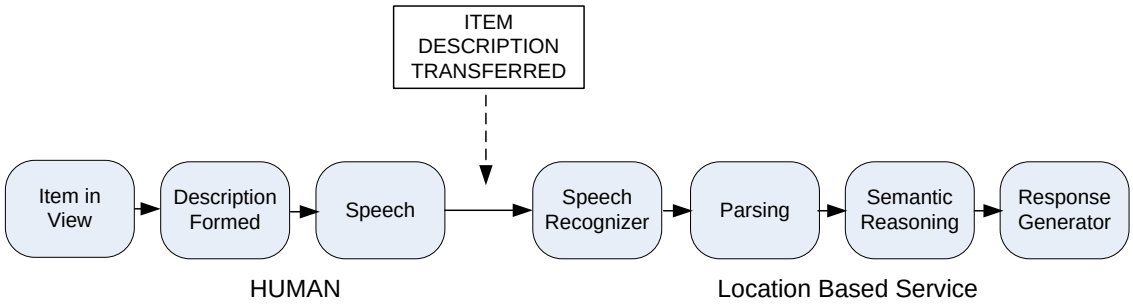


Figure 9.12: Comprehending a User Generated Description

Table 9-6 gives a number of examples of how the process may work using the fuzzy class approach. Each phrase is broken into attribute contributions, such as "small orange building" giving colour and size information. The differences between membership for all visible items and those attributes included in the description are compared to find the closest match. In this worked example the previous dataset is used again (from Table 9-2) and attributes not used are excluded from the calculation. In the first example "red building on the left" is converted into a high membership (1) for colour red, and a low value for left-right axis (0), indicating it is on the left, revealing the most likely match to be FOI 4.

In the next example "green FOI in the middle" does not result in a close match to any of the items in view, but FOI 5 is the most similar being a building in the middle. Finally "small orange building" returns FOI 3 as the most similar item to that description based on the exact colour match. These calculations assume an even weighting is assigned to each visual attribute, however depending on viewing conditions the attribute weighting may be adjusted, for example a foggy day may imposed difficulties in colour recognition, yet building position and size comprehension would remain constant. As an example by assigning foggy results with a 0.1 weighting for colour inputs, and 1.0 for size and position the 'small orange building' description returns a different result (FOI 4) which favours the building size attribute.

A comparison of the match strength gives a good indication of the confidence the system may have in its determination of the correct target feature. Where the confidence level is below a

defined threshold it would be possible to request additional attributes from the user, or to determine the most important clarifying characteristic for the candidates, as outlined in Section 9.5.2 and Table 9-5.

Table 9-6: Comprehending a Description by Determining Closest Matching Record

Phrase	Attribute				Viewing Conditions	
					Clear *1	Foggy *2
	Colour					
ID	(Red)	(Orange)	Size	Pos.	Total	Total
"Red building on the left."	1	0	null	0		
1	0.20	0.40	-	0.00	0.60	**0.06**
2	0.20	0.50	-	0.50	1.20	0.57
3	1.00	1.00	-	1.00	3.00	1.20
4	0.00	0.00	-	0.10	**0.10**	0.10
5	0.00	0.00	-	0.40	0.40	0.40
"Green FOI in the middle."	0.00	0.00	null	0.50		
1	0.80	0.40	-	0.50	1.70	0.62
2	0.80	0.50	-	0.00	1.30	**0.13**
3	0.00	1.00	-	0.50	1.50	0.60
4	1.00	0.00	-	0.40	1.40	0.50
5	1.00	0.00	-	0.10	**1.10**	0.20
"Small orange building."	0.00	1.00	0.00	null		
1	0.80	0.60	0.80	-	2.20	0.94
2	0.80	0.50	1.00	-	2.30	1.13
3	0.00	0.00	0.88	-	**0.88**	0.88
4	1.00	1.00	0.40	-	2.40	**0.60**
5	1.00	1.00	0.46	-	2.46	0.66

*1 Attribute weightings applied for clear viewing conditions are 1 (colour), 1(size), 1 (position).

*2 Attribute weightings applied for foggy viewing conditions are 0.1 (colour), 1(size), 1 (position).

9.7 Conclusions and Future Work

Natural language interfaces have the potential to offer mobile users superior interaction experiences, but to do so supporting models are required which translate data into a suitable form. This chapter develops one such conceptual model, a method for establishing the strongest combination of visual variables to be used when forming a referring expression for feature identification.

The model uses Bertin's visual variables as the foundation for forming the referring expression to describe the appearance of a FOI. The process begins by modelling the user's field

of view using a visual exposure model able to report key facts about the FOIs visibility, such as the façade area. The position and projective spatial relationships between visible FOIs are then calculated and combined with other known visual attributes to assign class memberships for a number of characteristics. Different combinations of these characteristics are compared to find the combination which leads to the least confusion with other items in view, determined by finding the solution which has the greatest distance to its nearest neighbour. The model is also useful in determining the most likely candidate based on parsing a user generated description.

The worked examples demonstrate that greater numbers of attribute combinations do not necessarily lead to stronger overall descriptions, as in Table 9-4 whereby the single *Colour* descriptor gave the strongest description. However the weighting between descriptors has not been investigated, nor the variations in weightings under different viewing conditions. For example the *colour* attribute may be deemed a less reliable indicator at dusk, or during rain, than *position* and therefore be awarded a lower contributing factor. Such variations and weightings should be the focus of future work.

A complete textured city model is required to run the model, which currently limits its use, however many community and corporate activities should satisfy this need in the near future. For example Google's "City in 3D program"(Google 2011), and sensor fusion techniques (El-Hakim *et al.* 1998, Pylvänäinen *et al.* 2010), however the timescale is unknown.

Views from high vantage points, where a large number of FOIs are visible, are not well supported by this model as only the most extreme of buildings will be identifiable. Also as it relies on an exhaustive search the computation is an NP-Hard task, which with a large number of attributes and for a large number of FOIs becomes computationally expensive (Dale and Reiter 1995). The model is most suited to urban corridors where the number of FOIs is more limited.

Other additions should include supporting counts in referring expressions, such as "The second white house from the left.", and superlatives (e.g. tallest). There also should be an evaluation of the description strength to determine if it is distinctive enough or additional clarification is required, perhaps through a two way dialogue to ensure the user has focussed on the correct item, or by adding descriptions from the surrounding buildings, for example "The red house next to the yellow house.". In addition foreground items can be included to assist in object location, such as "The white house behind the trees.".

Responding to a user's description has only been discussed from the point where variables have been identified, and no consideration has been given to the handling of ambiguous phrases or the use of unknown or complex variable structures.

Natural language interfaces assist users by closing the gap between human and machine, but a wide range of challenges still exist before this can become a reality.

Chapter 10: Conclusion

10.1 Summary of Research

The work presented in this thesis focused on the visual field, and how modelling a user's field of view can be of benefit to future LBS applications. A new visual exposure model was developed suitable for use in urban environments, able to generate a number of metrics to describe FOIs in a scene. An advance made was in accommodating vegetation within the exposure model, such that partial through canopy and clear under canopy views were modelled. The use of the model as part of a navigation task was also explored, such that a motorist is pre-informed of blind corners, or good overtaking opportunities on a defined route.

It is here considered that speech interfaces will become more widely adopted by LBS in the future, consequently there will be an increasing need to verbalise spatial information in a format that can be easily comprehended by the user. In support of this growing field the research presented in this thesis included the generation of referring expressions, whereby visible entities may be described from the user's perspective through a new model. The work presented here extends earlier research on this topic, in describing relationships in egocentric projective space, with a new model able to summarise relations in 3D space for visible buildings, with support for serialisation, storage and querying.

Finally, as part of the referring expression formulation, the issue of collecting colour information was examined. Here a new approach was developed and demonstrated that combined computer vision techniques with spatial analysis to automate the collection of house and boundary wall colours, filtering out foreground objects. This information is useful in conjunction with other datasets to form descriptions of the features in view, such that a LBS user's attention can be directed to a selected feature. In summary the concepts explored and developed included:

- A model to establish the visual exposure for urban features of interest, and a range of new visibility metrics (Chapters 2 & 3)
- The adaption of the visual exposure model to include partial through canopy and under canopy views (Chapter 4)
- The use of visibility modelling for a motorist's route, to identify overtaking opportunities and blind corners (Chapter 5)
- A new model of egocentric projective space that supports storage and querying, allowing relationships between features to be related to the user in an appropriate and meaningful way (Chapter 6)
- A new technique to combine computer vision and GIS datasets to automatically extract building colour description information, with filtering of foreground objects (Chapter 8)
- A method to determine the strongest referring expression dependent on the visibility of features (Chapters 7 & 9).

10.2 Summary of Outcomes and Conclusions

Visibility modelling in GIS has been concerned with establishing how much of the surrounding region is visible from a given location, but modelling how much of a feature is visible from the surroundings was generally ignored until Llobera (2003) developed a visual exposure model. This thesis has further developed visibility modelling capabilities by extending the concept of visual exposure modelling in urban areas, made possible by advances in data capture technology (i.e. LiDAR) which enable urban city models to be rapidly captured at high resolution. Such city models are required in visual exposure modelling so that building form and topography are captured, yet may be isolated from each other through the use of DTMs and DSMs (Chapters 3,4).

The modelling of views through vegetation was also incorporated in the visual exposure model by including a vegetation layer. This layer exists as a raster model which defines vegetation density and base canopy height at raster cell resolution. Vegetation information was sourced from multiband imagery in combination with a DSM, DTM and a ground survey. As a ray cast from the observer to a FOI passes through one of these "raster trees" it experiences an increased decay rate mimicking the partial blocking effect and loss of visibility of parts of the target feature. This is expressed in the results by decreasing the overall façade area visible, and in addition by the inclusion of a new metric which describes the distribution of vegetation across an FOI.

Such urban visibility models have many applications (e.g. planning, architecture), but here the mobile computing platform was chosen as the focus. LBS applications deliver information to users through confined interfaces in challenging environments, and new approaches to information filtering and presentation are required as the volume of data increases (Morville 2005). One of the most common LBS uses is in car navigation, and the role that visibility can play in establishing how far a driver can view the planned route ahead was developed in Chapter 5. The model was able to determine in advance the limited view entering corners, and the extended view on straight sections and hill tops.

Typically mobile users, especially car drivers, have restricted opportunities to view LBS screens and therefore speech interfaces are preferable for some tasks (e.g. navigational prompts). In addition as speech recognition accuracy improves it becomes a viable alternative in a wider range of LBS applications, including pedestrian based virtual city guides. However before a user can interact with an LBS through a speech interface, supporting models are required which can assist the user in locating surrounding features by the way they look and by relating their position in accordance with a user centred projective view. Such an egocentric projective spatial reasoning model was developed in Chapter 6 that summarises the relationships between two visible FOIs in the surrounding space. This model combines and extends previously published

models (Allen 1983, Billen and Clementini 2004, Billen and Clementini 2006, Clementini and Billen 2006), with the output from the visual exposure model developed here. The resulting outputs enable projective reasoning to be carried out in urban environments for FOIs in view, allowing the user to search for complex relationships in 3D space.

Spatial reasoning plays a big part in many parts of speech, and algorithms are required which allow a LBS application to determine what a user requests. An example of this was developed in Chapter 7 for the term "opposite", which relies on a projective relationship in turn serviced by a visual exposure model. These ascertain that both referenced FOIs are clearly in view, and that a third common feature may also be viewed between the FOIs.

Descriptions of features themselves are also useful in establishing, or clarifying, which building is the topic of discussion. To provide this functionality a model was developed in Chapter 9 which evaluates the most appropriate combination of visual attributes to be used in describing a FOI from the current scene. This uses a range of visual attributes, most of which could be sourced from GIS datasets, one found lacking was façade colour information and therefore a technique to automate the capture of this was developed in Chapter 8.

10.3 Current Limitations and Further Work

Beyond the limitations outlined in each chapter, there are a number of wider considerations which should be highlighted particularly regarding data sources, positioning, and speech recognition.

While LiDAR enables high resolution elevation models to be created in rapid time, the output is limited to a 2.5D representation of a city, leading to missing underpasses and building overhangs. This can be remedied to some degree by combining knowledge from other sources, as outlined in Chapter 4, but requires additional processing effort. In addition the model is typically in the order of 1 metre resolution thereby features with smaller footprints, such as telegraph poles and fences, are likely to be missing. Furthermore movable objects, such as cars and buses, are captured in the model and indistinguishable from permanent features. As the available spatial capture resolution capabilities increase so more of these small features will be included in the model, giving a greater need to determine which are fixed and which are movable. Chapter 5 presents a method to remove surface items from within the road boundary, but future models will also need to remove pedestrians and smaller surface objects from all regions of the model. This might be attempted by capturing multiple surface models over a short timeframe (e.g. a week) to determine the constant features between surface models. In the meantime the resolution limitations should be acknowledged, and the visual exposure model output should always be considered in light of these shortcomings.

The inclusion of a vegetation model, described in Chapter 4, improves modelled results in particular by maintaining the correct elevation for a user passing under canopy. However sourcing details for vegetation density and canopy base elevation on a city wide scale rapidly is currently a limiting issue. The solution looks to lie in full waveform LiDAR (Reitberger *et al.* 2006), but this has yet to be tested at city scale.

Estimating the error and uncertainty of results across the urban environment should be incorporated in future research. For example modelling visual exposure in an open city square will have a lower error compared to that modelled under bridges and through vegetation, and could be reflected in the generated referring expressions. Crowd sourcing may also provide valuable feedback as to the model's performance, and if widely adopted a system which allowed users to respond when a FOI indicated to be visible was in fact not, could lead to model improvements.

As the quantity and resolution of data sources increase so do the storage and processing requirements. Smartphones have become powerful mobile computers, but the task of processing live viewsheds is a computationally expensive one beyond their current real-time capabilities. A client-server approach satisfactorily provided a solution for experimental research purposes but with a greater number of clients so alternative server-sided approaches may be required which can scale up to support demand. Here cloud computing may provide a suitable answer, and this should be an area of future research.

It is well acknowledged that more research is required to deliver robust positioning solutions in urban environments. Unfortunately current mobile phone GPS chipsets do not support real-time positioning corrections from SBAS (e.g. WAAS), as is possible in dedicated GPS devices. However it is hoped that soon smartphone manufacturers will include such capabilities, and also carrier-phase enhancements (Wirola *et al.* 2007). Until this time computer vision image tracking (Botterill *et al.* 2010), foot tracking (Radoczky 2007) or a combination of both (Hide *et al.* 2010) probably provide the best interim urban solution.

This thesis has focussed on a framework to support speech based LBS applications, the next logical step would be to look at user trials and to expand the spatial natural language capabilities (i.e. NLG and NLP).

The future of Mobile Spatial Interaction (Fröhlich *et al.* 2009) is an exciting one, and as computers disappear from sight into everyday objects (Weiser 1993, Weiser 1994), so the role of information filtering and content delivery becomes increasingly important (Morville 2005). This is really just the beginning.

Appendix A:
Media Mapping: Using Georeferenced Images and Audio to provide supporting information for the Analysis of Environmental Sensor Datasets.

A.1 Summary

Field based environmental monitoring projects often fail to gather supporting temporal information on the surroundings, yet these external factors may play a significant part in understanding variations in the collected datasets. For example when sampling air quality the values may change as a result of a bus passing the sampling point, yet this temporal local information is difficult to capture at a consistently high resolution over extended time periods. Here we develop an application which runs on a mobile phone able to capture visual and audio data with corresponding time and location details. We also develop a desktop analysis tool which synchronises the display of this dataset with those captured from environmental sensors. The result is a tool able to assist researchers in understanding local changes in environmental datasets as a result of changes in the nearby surrounding environment.

A.2 Introduction

The analysis of temporal datasets in Geographic Information Systems (GIS) is often hampered by a lack of supporting relevant information on local conditions at the time of data capture. Being able to explain unpredictable variations in temporal datasets may depend on being able to understand the nature of the local environment at a very local scale. For example a passing vehicle may be the cause of a noted spike in airborne particulate matter, but unless this situational information is recorded the spike may never be explicitly explained. Sensor networks are able to supply background information revealing the wider situation, but a co-located synchronized set of sensors are required to understand the local situation during mobile data capture. Here we develop an application able to assist researchers in storing information on the local environment at the time of data capture.

The solution uses a mobile phone to store audio, visual, and location details against time such that during analysis the researchers are able to view the local environment at the time of data capture. A custom playback and analysis tool was also developed to combine this situational data with other time stamped data captured from the same location, allowing researchers fast access to the relevant contextual information during the analysis of the environmental data. In this

chapter we describe the function of the mobile and desktop applications, their design, and report on their usefulness in an air pollution monitoring study conducted in an urban area. The tools proved useful in explaining local spikes in air pollution data due to local events documented in the supporting images and audio data. The audio stream was also useful for allowing the researcher to take spatially and temporally attributed verbal notes in the field.

A.3 Background

Many studies have examined the relationship between air pollution and modes of urban commuting (Kingham *et al.* 1998, Van Roosbroeck *et al.* 2006, Gulliver and Briggs 2007, O'Donoghue *et al.* 2007, Briggs *et al.* 2008, Fruin *et al.* 2008). Air pollution data is highly temporal, changing across time and space, affected by global and local events. Local events, such as a bus passing the recording equipment, are hard to document and traditionally paper based records are kept. However manual references to such events are hard to integrate into any analysis, and sampling frequency is often inconsistent. Yet these factors are important as has been strongly argued by Briggs et al (2008) who state that local factors could be the cause of differences reported between studies including such things as "*building configuration, road layout, monitoring methods, averaging periods, season, meteorological conditions, vehicle, driving and walking behaviours, and the strength of in-vehicle sources*" (Briggs *et al.* 2008, 20).

To provide a better understanding of the surrounding environment at the time of data capture a second set of sensors can be used to automatically capture contextual information. This contextual information is not the primary data for research purposes, but a supporting dataset for the analysis phase. Technological advancements have made it possible to sense the environment more accurately, at higher sampling densities than ever before. Sensor networks may consist of electronic devices (Culler and Mulder 2004, Microsoft Corporation 2006), or citizens volunteering local environmental information via the internet (Goodchild 2007). En mass citizens may provide data without realizing it, such as the monitoring of mobile phones to estimate population movement (Ratti *et al.* 2006), or to estimate travel delays on motorway sections (Astarita *et al.* 2006).

The majority of electronic wireless sensor networks are static, distributed across a region at fixed sites, feeding information to a central facility which combines the data to build a picture of the surrounding conditions. For the purpose of mapping a commuter's exposure to air pollution the local level changes are also important, therefore a set of sensors should remain co-located with the air pollution sampling equipment. The data samples must also be at a high enough temporal resolution to record significant local events, and the data streams need to remain synchronized during data capture.

Mobile computing devices and smartphones have been proven useful in environmental monitoring enabling participants to collect and share data in real-time (Rudman *et al.* 2005). The MESSAGE consortium (Polak and Hoose 2008) have undertaken a number of projects using mobile phones as personal environmental sensors and data loggers. The mobile phones were equipped with a payload of environmental sensors able to record carbon monoxide, carbon dioxide, traffic volume, and nitrogen dioxide levels. Researchers exploring the city could feed data in real time to a data centre for processing, revealing current city wide air pollution trends. The individual trip data could also be replayed and mapped so that air pollution trails could be reviewed to visualize areas of poor air quality in the city. However a key aspect missing from this research was the facility to store video or image data from the user's surroundings. Therefore any analysis carried out at a later date would lack documented contextual detail. Although audio was used, it only provided an estimate of traffic volume and could not differentiate between vehicle types, or allow field researchers to take temporally and spatially attributed verbal notes.

Another research group developed a prototype system known as GeoMobSense (Kanjo *et al.* 2007). This toolkit allows private users to equip their own phones with the necessary facilities to log data from connected sensors. The phones themselves are used to display information, as well as log sound levels, while separate data loggers are used to record the environmental data. The resulting datasets can be exported and displayed on Google Earth, and other GIS applications. Again this toolkit fails to store continuous image sequences, or to save a spatially attributed audio file. Therefore any post-capture analysis is hampered by a lack of documentation on the surrounding situation during the field study.

Multimedia files provide a useful companion dataset to prompt recall of events which occurred during data capture. They provide an extra channel of information useful when linked to GISs (Cartwright *et al.* 2007), however the video, image, or sound files are normally linked to a point, as in Media Mapper (Red Hen Systems 2009). This creates a one way relationship, only allowing the corresponding multimedia clip to be found when the user clicks on a map location, essentially using the map as document retrieval interface. The content of the multimedia file itself is not spatially attributed, and the ability to jump to the corresponding position in the video file for a given map location has to be performed manually.

There have been a number of attempts to more closely link the multimedia content to space through dynamically geo-referencing multimedia files. Spatial information is encoded into the file through an appropriate technique such that at any point in the video or audio the corresponding location may be referenced. For example specialist equipment can turn GPS location information into audio data, in a similar way that a modem is able to turn computer data into audio to send it across a telephone line, which can then be recorded to the audio track alongside the video data. An example of this technology is CamNav Mapper (Blueglen Ltd

2009). This allows a user to search through a video file, and at all times be able to display the corresponding recording location in a GIS. However the connection is uni-directional, meaning the video is able to provide location information to the GIS, but it is not possible to initiate a search for the corresponding part in the video from the GIS.

Jaejun (2002) developed an application which supports bi-directional searching, permitting the user to find relevant video information from selecting a GIS location, or for finding the filming location by searching the video file. Similarly Zeiner (2005)developed an application able to fuse GPS location and video using timestamp information collected from synchronized clocks. For still images a set of points are created in the GIS, however for video recorded while moving a track denotes which parts of the video correspond to which geographic location. They also explore the use of data standards in providing geo-multimedia tools via the World Wide Web, with particular focus on the overlap between web mapping standards, metadata standards, and video streaming standards. These tools are not however designed with the ability to integrate other temporal datasets such as required for environmental research.

Other studies have used or developed analysis tools to visualize environmental datasets with local situational data (Arnold *et al.* 2004, Kaur *et al.* 2006, Terwoert 2009). However while these often include the ability to link photographic images with environmental data, they appear to lack a tightly integrated mapping facility.

For our research the requirement was for an application which could provide a high level of integration between temporal multimedia and location datasets, with the ability to support additional datasets collected from synchronized sensors. The capture device needed to be small, lightweight, and mobile such that it could be carried by a pedestrian or cyclist easily for extended periods of time. The datasets for location, audio and image needed to be tightly coupled such that no synchronisation issues could occur during long field trials. The analysis tools needed to be easily operated by an untrained GIS user, such that they could search through the datasets to interact with any of the captured data streams while maintain sync with the other linked data sources. We therefore looked at developing the data logging tool on a GPS equipped smartphone, which is a highly portable programmable device available at low cost. As a result of using a single programmable device the GPS, audio, and image datasets are tightly coupled, removing the need to synchronize clocks, and guaranteeing data streams remain in sync indefinitely. Additionally the multimedia files are georeferenced at the time of data capture removing the need for any post-capture data processing. In contrast to other applications which use laptops to capture and process media, the smartphone approach offers a robust, small, and very portable platform which may be easily carried by pedestrians. Finally our analysis tool supports a tri-directional search mechanism, such that users may drive the search by moving through the audio media, mapping interface, or by interrogating charts of the additional sensor data. This means the user is able to easily capture and analyze data using any of three mediums

(location, time, graph value). In the next section we look in more detail at the applications developed during this research.

A.4 Application Development

In this study we developed two applications to assist in the process of recording the surrounding situational conditions. The first application runs at the time of data capture in the field on a smartphone equipped with Assisted GPS (A-GPS), and stores both an audio and visual record of the surroundings. A-GPS is particularly useful for urban based research as it provides a faster start-up location solution throughout a greater range of urban environments, such that a position could be found more quickly and maintained more consistently. Furthermore the phone selected for this research had a high sensitivity GPS chipset, enabling locations to be calculated across a high proportion of the city, including inside some single storey buildings. The second application developed for this research runs on a desktop computer and assembles independent data streams against a common timeline, such that the user may easily browse through multiple datasets whilst maintaining sync between them, at all times being able to refer to the corresponding situational image and audio data. We discuss each application in more detail in Sections 3.1 and 3.2, starting with the mobile data capture tools.

A.5 Mobile Data Capture Application

The main design criterion for the data capture device was that it would be used in urban studies everyday over an extended period of a few months. It therefore had to be small, robust, light weight, offer a large data storage capacity, be able to capture audio, imagery, and run on battery for at least 90 minutes to ensure an entire urban commute could be captured in a single session. We decided to use a Nokia N82 smartphone to carry out these tasks as they can be programmed easily using the Python language, incorporate a high sensitivity A-GPS able to function adequately in urban canyons, and have a high quality camera. Furthermore they are able to use micro-SD cards for data storage, are smaller than any laptop or netbook computer, and have good battery life.

Nokia Series 60 smartphones can be programmed in three main languages, which are C, Java, and Python. Python is very suitable for rapid development and allows the developer to access core phone hardware through supported Application Program Interfaces (APIs). The hardware access required for this project included GPS hardware, audio, and screen display. Our initial application was designed to record a continuous video and audio feed to the micro-SD card at 15 frames per second, while logging GPS locations every second. The phone supports the ability for Python applications to request the position of the current playhead in the video file during

recording, enabling the application to log the GPS position information along with the current video position to ensure a tight coupling of the location and video datastreams. The data capture application performed well, and as it used MP4 compression a full one hour video with GPS log files occupied only around 70MB. However the continuous video capture depleted a fully charged battery in 60 minutes, and made the mobile phone run fairly hot. After discussions with the air pollution research team we looked at an alternative solution to record still images at regular intervals.

The next iteration of our application, and the one used in field trials, records an audio file continuously but captures still images at the rate of 1 image every 3 seconds. The audio file is recorded to a WAV file at 8kHz, again the playhead position is stored with each GPS update such that for every image the location and position in the audio file is known explicitly.

To link the audio and image files with GPS a common timeline primary key is required. GPS time was considered unsuitable for the base timeline as the user may lose the GPS signals when moving inside buildings. Therefore the Python time from the phone clock was considered more reliable, and forms the baseline to which all other datasets are synced. GPS time is however stored as an additional attribute in the log file in case it is required later.

Specialist sensors were used to sample the air quality, recording the concentration of particulate matter at various sizes (PM_{10}, $PM_{2.5}$ and PM_1), ultrafine particles (UFP), temperature and carbon monoxide levels. These devices all have internal clocks which were synced to the nearest second to the clock on the mobile phone before each journey. During transit the mobile phone was mounted facing forwards on either the car dashboard, bike handlebars, or the strap of a rucksack to ensure the camera could capture a clear view of the oncoming route.

Figure A.1 shows the data collected with each journey. The smartphone tags each image with a unique identification number based on the Python time in seconds, ensuring that images correspond to a single log entry. Log entries record the position of the playhead in the sound file, Python time, GPS time, and the GPS latitude and longitude. Additionally we stored the cell tower identification value so that approximate locations can be determined if GPS positioning is lost. As well as recording GPS position we also record speed, GPS accuracy, heading, the number of satellites visible, and number used for the position solution. This enables us to carry out analysis on the location accuracy if required at a later date.

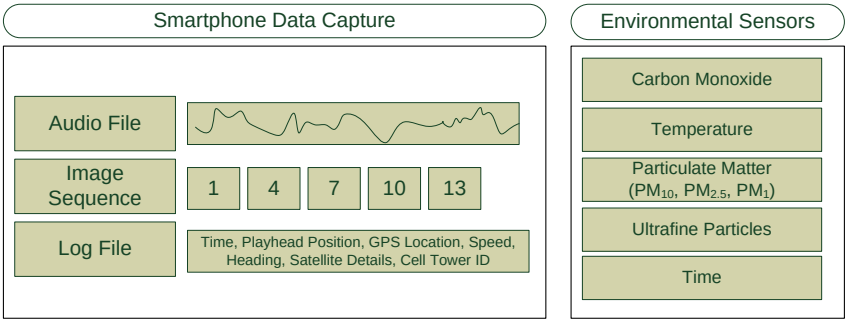


Figure A.1: An Overview of the Data Captured

A.6 Data Analysis and Playback

To be able to efficiently analyse the large volume of time series data collected it was necessary to build a custom application which allowed non-GIS users the ability to review the data easily from a single interface. To do this we wrote a custom application using C# .NET which made use of a number of open source libraries for charting, map display, and coordinate projection (Figure A.2). One of the key criteria in our application design was that all the datasets should be synced and remain in sync across all dataset viewers while the user explored the data. For example selecting a map point should display the relevant air pollution data, the corresponding street image, and move the sound file playback to the correct location so any relevant audio notes and background noises could be heard. In the following section we discuss each of these data visualisation elements, and describe the methods through which the datasets are kept in sync.

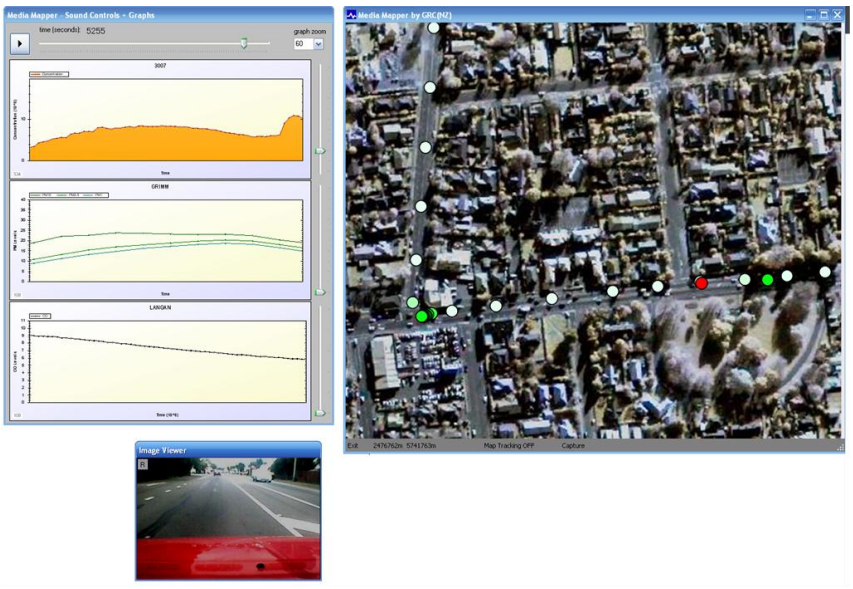


Figure A.2: Data Analysis and Playback Tool

A.6.1 Mapping

The mapping element makes use of the Piccolo Zoomable User Interface (ZUI) graphics framework. This open source library is not specifically designed for map display but offers powerful functions enabling rapid development of ZUIs. Piccolo is able to create scene graphs consisting of both vector and raster nodes, which may be easily animated to change location, shape or colour. It also provides the functionality to smoothly zoom and pan around the data, capturing mouse and keyboard input. Events can be assigned to graphical objects such that a user may interact with map items and trigger a custom action. In our application we use this functionality to link map objects to custom search functions, so for example clicking on a GPS track point moves the charting tool along to display the air quality values at that location, and moves the audio playhead forwards or backwards to the corresponding audio sample.

The mobile phone logs GPS data using the WGS84 coordinate system, however as Piccolo is unaware of geographic coordinate systems it requires projected datasets. We used New Zealand Map Grid (NZMG) as our projected coordinate system for all datasets. The background mapping layers included Quickbird satellite imagery, and Land Information New Zealand road centrelines transformed using GRASS to NZMG. The vector datasets, typically ESRI Shapefiles, were converted to a BNA text format using OGR libraries before being loaded and displayed by Piccolo, while the raster datasets (JPEG) were natively supported. ESRI World files were required with each raster image to provide pixel resolution and location coordinates values, so that raster datasets could be loaded into the correct position.

The GPS data, once downloaded from the mobile application to the desktop application, is transformed into the NZMG coordinate system using the most accurate NTv2 projection transformations provided by the Proj4 libraries. These points were then displayed as Piccolo vector nodes over the base mapping, each node with a hidden tag holding the corresponding capture time primary key. This tag corresponds to the phone capture time in seconds, and is also recorded in the log file against the playhead position and location details. These tag data enable the software to instantly locate corresponding relevant information in other datasets when a user clicks on a map object ensuring the system is very responsive.

A.6.2 Audio

The audio playback is handled using Microsoft DirectSound libraries, which enable the rapid development of software able to control audio datasets. Here we use basic playback control features (i.e. play, pause, forward, rewind), and the functionality to read the current playhead position during payback. From this we can calculate the current play time in seconds from the start of the sampling period, and therefore find the corresponding log entries which hold the

geographic location and relevant image filenames. As the audio file is captured at 8000 samples per second it has the highest resolution of any of the datasets. Therefore any searches performed from the user cueing or reviewing this dataset require an additional step to find the most relevant (i.e. closest) timestamp in the log files. This was performed by simply ranking the difference in time from the audio position to all log entries, the first item being selected as the nearest. This functionality allows the user to review the audio file and also see corresponding imagery, location and chart data.

Due to memory issues when loading a 90 minute long audio file, a buffered playback method was required. Only a small section of the file is loaded into a memory buffer, and this buffer is constantly filled from the disk file as audio playback progresses. The performance impacts of this technique were minimal, and rapid audio reviewing and cueing are still possible.

A.6.3 Street Images

The smartphone captures an image every 3 seconds in JPEG format, and labels it with the appropriate timestamp. The log file holds a list of these timestamp filenames along with their corresponding location and other GPS details. When the user selects a location on the map the tag (with each Piccolo node object) holds the timestamp details, and therefore the corresponding image can instantly be loaded without performing any search other than for the filename in the file system. When the user controls the audio playback the nearest timestamp information is found in the log file, and from this the corresponding picture name can be generated. Similarly when the user moves through the chart information the nearest timestamp in the log file is found and used to determine the appropriate image to display.

The image display is supported using native .NET libraries, with functionality to rotate the image sequence which is useful if the phone has been placed on its side during image capture.

A.6.4 Charting

The environmental sensor datasets are charted using the open source ZedGraph libraries. These provide sophisticated charting capabilities, and allow the programmer to link into many key events such as when the user pans across the chart, clicks in the charting area, or changes the scale on a chart axis. In our case the x-axis was allocated to time in seconds since the start of sampling. Ideally the smartphone would be turned on first to ensure the audio file timeline starts before the environmental data, but negative time values are also supported. The y-axis displays the values from the relevant sensor. The display automatically scales the y-axis according to the

values in the entire loaded dataset, although manual scaling is also possible to zoom into values for smaller sample periods.

The ZedGraph library was used to create line graphs for each of the environmental sensor types. These can remain static to allow the researcher to view the entire datasets while replaying map, image and audio datasets. Alternatively the graphs can be dynamically linked to the playback such that they pan along the x-axis (time) automatically as the data log is replayed. In this case the current playback time is shown on the far left of the chart.

If the user clicks within the chart area the application retrieves the corresponding time (x-axis value) and moves the playback position to that value. This allows the researcher to instantly find the current geographical location for any spikes noted in the environmental datasets. Also as the audio and image files remain in sync the researcher is also able to look and listen to information from the surroundings at that point.

For reporting purposes the system supports high quality output of the graphs, by simply double clicking on them. There is also functionality to allow the user to export snapshots of any interesting results, effectively using this tool to produce a filtered dataset. To do this they simply click a button during playback and the GPS, picture link reference, time and date, and graph data are exported to a text file. As each environmental sensor operates at a different sampling frequency the application interpolates values between readings (i.e. straight line between known values). When the data is output the interpolated value is used if an actual reading value is not present for the current playback position.

A.7 User Interaction

One of the key differences in the analysis application developed for this research to those reviewed earlier, is the ability to initiate a search from any of three linked interfaces. Figure A.3 summarises the processes required to provide this functionality.

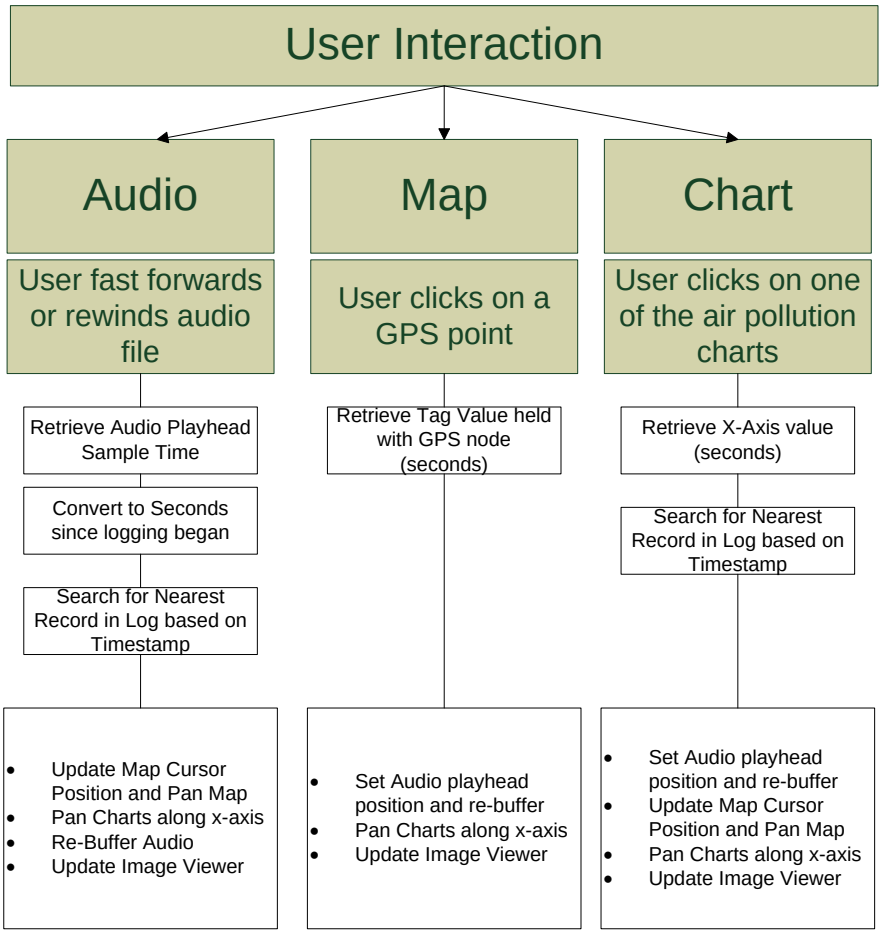


Figure A.3: User Interaction via Audio, Map, or Chart Interfaces

During normal playback the system carries out an audio update event every 4000 samples (500ms). This ensures that the displays of each dataset remain in sync, without impacting the performance on slower computers.

In the next section we look at a number of examples which demonstrate the usefulness of this application

A.8 Examples

In the following section we look at data collected in air pollution studies from Christchurch, New Zealand. The figures illustrate how useful the contextual information was in explicitly explaining peaks in the air quality datasets. In Figure A.4 we can see that the air quality spikes in the top left graph occur just after the bus (pictured) pulls away from the field observer. In addition the location information and map allow the researchers to easily and quickly identify where in the city these interesting results occurred.

Figure A.4: Analysis of Peak Resulting from Bus Pulling Away

In the next example the air quality levels are much better while exploring a pedestrian precinct, with low particulate matter concentrations (Figure A.5). However following this a sharp rise in particulate matter can be noted. By clicking on the spike in the graph display area the research is able to see, from the map location and supporting images, that this spike correlates to when the field observer entered a multi-storey car park.

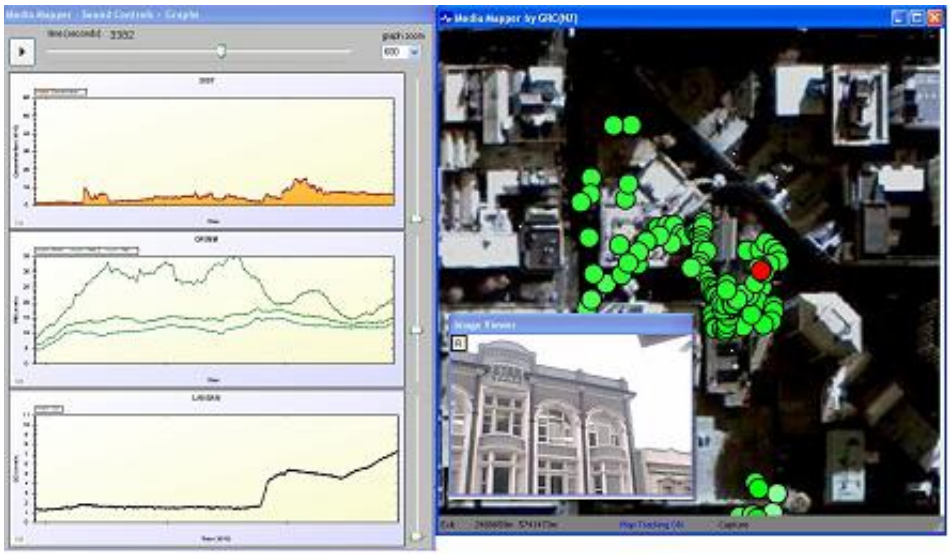


Figure A.5: Analysis of Air Data while Walking in Pedestrian Precinct

In the final example data were collected while travelling on a bus around the city. A series of spikes can be noted in the carbon monoxide levels at fairly regularly spaced intervals, as shown in Figure A.6. By reviewing the audio data, just before each spike, it was possible to hear the sounds of the door opening, ticket machine being operated and so on, and therefore conclude that these spikes may be related to door opening events. It is also worth noting that a camera facing

forwards would not have picked up this information, thus demonstrating the value of collecting audio data.

Figure A.6: Bus Trials of Air Pollution Monitoring Equipment

A.9 Conclusion

In this research we have demonstrated that a smartphone may be used as a suitable data capture tool, able to accurately and consistently capture location attributed audio and visual data while operating in an urban space. The onboard Assisted-GPS (A-GPS) was able to rapidly locate the user when the device was first turned on, and maintain position throughout exploration of outdoor urban areas. The audio and image quality was suitable for later analysis to identify key events which may be associated with changes in the local environment as a result of short temporal events (such as a bus pulling out) or changes in the nature of the physical environment (such as entering a car park). Python provided an excellent programming language for rapid application development on mobile phone, with the necessary functionality to use the phone's hardware such that explicit geo-referenced image and audio could be processed on board the phone, rather than during later post-processing in the desktop environment. This ensures that data sync between audio, location, and imagery datasets may be maintained indefinitely. The smartphone platform also proved to be rugged enough for daily trials over a period of several months, and had battery and storage facility to cater for long urban commutes (tested up to 120 minutes).

We also demonstrated that a simple desktop application able to maintain sync between mapping, environmental, audio, and visual datasets proved useful in the analysis phase of the

research. The tool allowed non-GIS researchers the functionality to easily explore environmental datasets while maintaining links to the corresponding contextual information.

The project was completed using a number of open source tools and libraries, without which the development would not have been possible in the allotted time or within budget. Future versions of the application should include the ability to store bookmarks against time, allowing the user to add keywords or notes, which would be particularly useful when revisiting previous datasets, or to store comments for other project collaborators to view. More powerful query tools would be useful too, such that map points may be highlighted or hidden based on the closest environmental data. Finally the ability to search audio files for speech would be useful in longer sampling runs such that the system could automatically identify where audio notes have been taken.

A.10 Acknowledgements

The authors would like to acknowledge the contributions made by Kreepa Shrestha, and Woodrow Pattinson who carried out the extensive field trials. Also Justin Harrison for setting up and supporting the environmental sensor equipment. This research would not have been possible without funding support from the Geospatial Research Centre (NZ) and New Zealand Transport Agency.

References

Abella, A. & Kender, J. R., (1993) Qualitatively describing objects using spatial prepositions. *In* Kender, J. R. (ed.) *Proceedings of the eleventh national conference on Artificial intelligence.* Washington, D.C., AAAI Press.

Abowd, G. D., Atkeson, C. G., Hong, J., Long, S., Kooper, R. & Pinkerton, M. (1997). Cyberguide: A mobile context-aware tour guide. *Wireless Networks, 3*(5), 421-433.

AgiSoft. (2010). PhotoScan software. Retrieved 11 October 2010, from http://www.agisoft.ru/products/photoscan/

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM, 26*, 832-843.

Allen, J. F. (2003) Natural language processing. *In* Ralston, A., Reilly, E. D. & Hemmendinger, D. (eds.) *Encyclopedia of Computer Science.* John Wiley and Sons Ltd.

Allport, D., Antonis, B. & Reynolds, P. (1972). On the division of attention: a disproof of the single channel hypothesis. *Quarterly Journal of Experimental Psychology, 24*, 225-235.

Amri Musliman, I., Alizadehashrafi, B., Chen, T. K. & Abdul-Rahman, A. (2010) Modelling visibility through visual landmarks in 3D navigation using geo-DBMS. *In* Neutens, T. & De Maeyer, P. (eds.) *Developments in 3D Geo-Information Sciences.* Springer.

Arnold, S. J., ApSimon, H., Barlow, J., Belcher, S., Bell, M., Boddy, J. W., Britter, R., Cheng, H., Clark, R. & Colvile, R. N. (2004). Introduction to the DAPPLE Air Pollution Project. *Science of the Total Environment, 332*(1-3), 139-153.

Astarita, V., Bertini, R. L., d'Elia, S. & Guido, G. (2006). Motorway traffic parameter estimation from mobile phone counts. *European Journal of Operational Research, 175*(3), 1435-1446.

Azuma, R. T. (1995). A Survey of Augmented Reality. *Teleoperators and Virtual Environments, 6*(4), 355-385.

Baer, W., Baer, N., Powell, W. & Zografos, J., (2005) Advances in terrain augmented geometric pairing algorithms for operational test. *ITEA Modelling and Simulation Workshop.* Las Cruces, NM.

Bartie, P. & Mackaness, W. A. (2006). Development of a speech-based augmented reality system to support exploration of cityscape. *Transactions in GIS, 10*(1), 63-86.

Bartie, P., Mills, S. & Kingham, S. (2008) An egocentric urban viewshed: A method for landmark visibility mapping for pedestrian location based services. *In* Moore, A. & Drecki, I. (eds.) *Geospatial Vision - New Dimensions in Cartography.* New Zealand, Springer.

Bartie, P., Reitsma, F., Kingham, S. & Mills, S. (2010). Advancing visibility modelling algorithms for urban environments. *Computers Environment and Urban Systems, 34*(6), 518-531.

Bartie, P., Reitsma, F. & Mills, S. (2011). Building colour terms: A combined GIS and stereo vision approach to identifying building pixels in images to determine appropriate colour terms *Journal of Spatial Information Science, 2*, 59-83.

Bartie, P., Clementini, E., Reitsma, F. & Kingham, S. (forthcoming). A qualitative model for describing the arrangement of visible cityscape objects from an egocentric viewpoint. *GeoInformatica*.

Bay, H., Ess, A., Tuytelaars, T. & Gool, L. V. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding, 110*(3), 346–359.

Beeharee, A. & Steed, A., (2005) Filtering location-based information using visibility. *Location and Context Awareness.* Germany, 306-315.

Benavente, R., Vanrell, M. & Baldrich, R. (2006). A data set for fuzzy colour naming. *Color Research & Application, 31*(1), 48-56.

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. *Environment and Planning B, 6*(1), 47-65.

Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M. & Row-Farr, J. (2006). Can you see me now? *ACM Transactions on Computer-Human Interaction (TOCHI), 13*(1), 100-133.

Berlin, B. & Kay, P. (1969). *Basic color terms: Their universality and evolution*: Univ. of California Press, Berkeley.

Bertin, J. (1983). *Semiology of graphics*: Madison: University of Wisconsin Press.

Betke, M. & Nguyen, H., (1998) Highway scene analysis from a moving vehicle under reduced visibility conditions. *Proc. of the International Conference on Intelligent Vehicles.* Stuttgart, Germany, Citeseer, 131–136.

Billen, R. & Clementini, E., (2004) Introducing a reasoning system based on ternary projective relations. *Developments in Spatial Data Handling.* Leicester, England 381-394.

Billen, R. & Clementini, E., (2005) Semantics of collinearity among regions. *In* Meersman, R., Tari, Z. & Herrero, P. (eds.) *On the Move to Meaningful Internet Systems 2005: OTM Workshops - 1st Int Workshop on Semantic-based Geographical Information Systems (SeBGIS'05).* Agia Napa, Cyprus, Springer-Verlag, 1066-1076.

Billen, R. & Clementini, E. (2006). Projective relations in a 3D environment. *Lecture Notes in Computer Science, 4197*, 18.

Birchfield, S. & Tomasi, C. (1999). Depth discontinuities by pixel-to-pixel stereo. *International Journal of Computer Vision, 35*(3), 269-293.

Black, N. & Coast, S. (2007). Geodata collection in the 21st century. *Bulletin of the Society of University Cartographers, 41*(1-2), 3-8.

Blueglen Ltd. (2009). CamNavMapper [Electronic Version] from http://www.blueglen.com/prod_camnav_single.htm.

Boring, E. (1964). Size constancy in a picture. *American Journal of Psychology, 77*, 494-498.

Borrmann, A. & Rank, E. (2009). Topological analysis of 3D building models using a spatial query language. *Advanced Engineering Informatics, 23*(4), 370-385.

Botterill, T., Mills, S. & Green, R. (2010). Bag of words driven, single camera simultaneous localization and mapping. *Journal of Field Robotics, 28*(2), 204-226.

Boynton, R. M. & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research, 30*(9), 1311-1317.

Bradski, G. & Kaehler, A. (2008). *Learning OpenCV*, Sebastopol, CA: O'Reilly.

Brenner, C., Haala, N. & Fritsch, D. (2001). Towards fully automated 3D city model generation. *Automatic Extraction of Man-Made Objects from Aerial and Space Images III*, 47–57.

Briggs, D. J., de Hoogh, K., Morris, C. & Gulliver, J. (2008). Effects of travel mode on exposures to particulate air pollution. *Environment International, 34*(1), 12-22.

Brown, B., Wright, H. & Brown, C. (1997). A post-occupancy evaluation of wayfinding in a pediatric hospital: research, findings and implications for instruction. *Journal of Architectural and Planning Research, 14*(1), 35-51.

Bruin, J. d., Fensel, D., Kerrigan, M., Keller, U., Lausen, H. & Scicluna, J. (2008). *Modeling Semantic Web Services: The Web Service Modeling Language.* : Springer.

C3 Technologies AB. (2009). Stockholm City [Electronic Version], 2010. Retrieved 16 August 2010 from http://cooper.c3technologies.com/demo/myvr/sthc.html.

Caduff, D. & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing, 9*(4), 249-267.

Cahill, V., Gray, E., Seigneur, J. M., Jensen, C. D., Chen, Y., Shand, B., Dimmock, N., Twigg, A., Bacon, J., English, C., Wagealla, W., Terzis, S., Nixon, P., Di Marzo Serugendo, G., Bryce, C., Carbone, M., Krukow, K. & Nielsen, M. (2003). Using trust for secure collaboration in uncertain environments. *IEEE Pervasive Computing, 2*(3), 52-61.

Caldwell, D., Mineter, M., Dowers, S. & Gittings, B., (2003) Analysis and visualisation of visibility surfaces. *GeoComputation.* 8 - 10 September 2003, University of Southampton,UK, GeoComputation CD-ROM.

Canming. (2010). Emgu.CV. Retrieved 15 September 2010, from http://www.emgu.com/wiki/index.php/Main_Page

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Analysis and Machine Intelligence, 8*(6), 679-698.

Carswell, J. D., Gardiner, K. & Yin, J. (2010). Mobile visibility querying for LBS. *Transactions in GIS, 14*(6), 791-809.

Cartwright, W., Peterson, M. P. & Gartner, G. F. (2007). *Multimedia cartography*: Springer Verlag.

Chen, M. Y., Sohn, T., Chmelev, D., Haehnel, D., Hightower, J., Hughes, J., LaMarca, A., Potter, F., Smith, L. & Varshavsky, A. (2006). Practical metropolitan-scale positioning for GSM phones. *Lecture Notes in Computer Science, 4206*, 225-242.

Cheverst, K., Davies, N., Mitchell, K., Friday, A. & Efstratiou, C., (2000) Developing a context-aware electronic tourist guide: some issues and experiences. *Conference on Human Factors in Computing Systems.* Netherlands, ACM, 17-24.

Chincholle, D., Goldstein, M., Nyberg, M. & Eriksson, M. (2002). Lost or Found? A usability evaluation of a mobile navigation and location-based service. *Mobile HCI, 2411*, 211-224.

Cho, D. J. (2004). An assisted GPS acquisition method using L2 civil signal in weak signal environment. *Journal of GPS, 3*(1-2), 25-31.

Chuang, J., Stone, M. & Hanrahan, P. (2008). A probabilistic model of the categorical association between colors. *Color Imaging Conference* Retrieved 17 July 2010, from http://vis.stanford.edu/files/2008-ColorNames-CIC-Talk.pdf

Clementini, E. & Di Felice, P. (1995). A comparison of methods for representing topological relationships. *Information Sciences-Applications, 3*(3), 149-178.

Clementini, E. & Di Felice, P. (1997). Approximate topological relations. *International Journal of Approximate Reasoning, 16*(2), 173-204.

Clementini, E., Felice, P. D. & Hernández, D. (1997). Qualitative representation of positional information. *Artificial Intelligence, 95*(2), 317-356.

Clementini, E. & Billen, R. (2006). Modeling and computing ternary projective relations between regions. *IEEE Transactions on Knowledge and Data Engineering, 18*, 799-814.

Clementini, E. (2009). *A Conceptual framework for modelling spatial relations.* Institut National des Sciences Appliquées de Lyon, Lyon.

Cohn, A. G. (1997). Qualitative spatial representation and reasoning techniques. *Lecture Notes In Computer Science, 1303*, 1-30.

Cohn, A. G. & Hazarika, S. M. (2001). Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae, 46*(1-2), 1-29.

Cooper, G. R. J. (2005). Analysing potential field data using visibility. *Computers and Geosciences, 31*(7), 877-881.

Culler, D. E. & Mulder, H. (2004). Smart sensors to network the world. *Scientific American, 290*(6), 84-91.

Dale, R., Mellish, C. S. & Zock, M. (1990). *Current research in natural language generation*: Academic Press.

Dale, R. & Reiter, E. (1995). Computational interpretations of the Gricean maxims in the generation of referring expressions. *Cognitive Science, 19*(2), 233-263.

Davidson, D. A., Watson, A. I. & Selman, P. H. (1993) An evaluation of GIS as an aid to the planning of proposed developments in rural areas. *In* Mather, P. M. (ed.) *Geographical Information Handling: Research and Applications.* London, Wiley.

Davies, N., Cheverst, K., Mitchell, K. & Friday, A., (1999) Caches in the air: disseminating tourist information in the Guide system. *Second IEEE Workshop on Mobile Computing Systems and Applications.* 25-26th February 1999, New Orleans, Louisiana, IEEE, 11-19.

De Floriani, L. & Magillo, P. (1994). Visibility algorithms on triangulated digital terrain models. *International Journal of Geographical Information Systems, 8*(1), 13-41.

De Floriani, L., Marzano, P. & Puppo, E. (1994a). Line-of-sight communication on terrain models. *International Journal of Geographical Information Systems, 8*(4), 329-342.

De Floriani, L., Montani, C. & Scopigno, R. (1994b). Parallelizing visibility computations on triangulated terrains. *International Journal of Geographical Information Systems, 8*(6), 515-531.

De Floriani, L., Magillo, P. & Puppo, E., (1997) Building and traversing a surface at variable resolution. *Proceedings of the IEEE Visualization Conference.* Phoenix, AZ, USA, IEEE Comp Soc, 103-109.

De Floriani, L., Magillo, P. & Puppo, E. (2000). VARIANT: A system for Terrain modeling at variable resolution. *GeoInformatica, 4*(3), 287-315.

De Floriani, L. & Magillo, P. (2003). Algorithms for visibility computation on terrains: A survey. *Environment and Planning B: Planning and Design, 30*(5), 709-728.

De Smith, M. J., Goodchild, M. F. & Longley, P. (2007). *Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools*: Troubador Publishing.

Dean, D. J. (1997). Improving the accuracy of forest viewsheds using triangulated networks and the visual permeability method. *Canadian Journal of Forest Research, 27*(7), 969-977.

Dillemuth, J. (2005). Map design evaluation for mobile display. *Cartography and Geographic Information Science, 32*(4), 285-301.

Douglas, K. v. D., James, A. L. & Jason, I. H. (2006). *The Design of Sites: Patterns for Creating Winning Web Sites (2nd Edition)*: Prentice Hall PTR.

Du, H. & Crestani, F., (2004) Spoken versus written queries for mobile information access. *Mobile and Ubiquitous Information Access.* Udine, Italy, Springer, 67-78.

Duckham, M. & Kulik, L., (2006) Location privacy and location-aware computing. *In* Drummond, J. (ed.) *Dynamic & mobile GIS: investigating change in space and time.* Boca Raton, FL, CRC Press, 34–51.

Duckham, M., Winter, S. & Robinson, M. (2010). Including landmarks in routing instructions. *Journal of Location-Based Services, 4* (1), 28-52.

Egenhofer, M. & Mark, D., (1995) Naive Geography. *In* Frank, A. & Kuhn, W. (eds.) *COSIT.* Austria, Springer-Verlag, 1-15.

Egenhofer, M. J. & Herring, J., (1990) A mathematical framework for the definition of topological relationships. *Fourth International Symposium on Spatial Data Handling.* Switzerland, 803-813.

Egenhofer, M. J., (1991) Reasoning about binary topological relations. *Second Symposium on Large Spatial Databases.* Switzerland, Springer-Verlag, 143-160.

Egenhofer, M. J. & Franzosa, R. D. (1995). On the equivalence of topological relations. *International Journal of Geographical Information Systems, 9*(2), 133-152.

El-Hakim, S. F., Brenner, C. & Roth, G. (1998). A multi-sensor approach to creating accurate virtual environments. *Journal of Photogrammetry and Remote Sensing, 53*, 379-391.

Elias, B. (2003a) Extracting landmarks with data mining methods. *In* Kuhn, W., Worboys, M. F. & Timpf, S. (eds.) *Spatial information theory.* Berlin, Springer.

Elias, B., (2003b) Determination of landmarks and reliability criteria for landmarks. *5th Workshop on Progress in Automated Map Generalization.* Paris, France.

Elias, B. & Brenner, C. (2004) Automatic generation and application of landmarks in navigation data sets. *In* Fisher, P. F. (ed.) *Developments in Spatial Data Handling.* Berlin, Springer.

Ennesser, F. & Medioni, G. (1995). Finding waldo, or focus of attention using local color information. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 17*(8), 805-809.

Ertan, S., Lee, C., Willets, A., Tan, H. & Pentland, A., (1998) A wearable haptic navigation guidance system. *Digest of the Second International Symposium on Wearable Computers.* Pittsburgh, 164-165.

Espinoza, F., Peterson, F., Sandin, P., Nystrom, H., Cacciatore, E. & Bylund, M., (2001) GeoNotes: Social and navigational aspects of location-based information systems. *In* Shafer, S. (ed.) *Ubicom 2001.* September 30 - October 2, Atlanta, Georgia, Springer, 2-17.

Fantino, M., Mulassano, P., Dovis, F. & Lo Presti, L. (2008). Performance of the proposed Galileo CBOC modulation in heavy multipath environment. *Wireless Personal Communications, 44*(3), 323-339.

Feiner, S., Macintyre, B., Seligmann, D. (1993). Knowledge based augmented reality. *Communication of the ACM, 36*(7), 53-62.

Finlayson, G., Hordley, S. & Drew, M., (2002a) Removing shadows from images using retinex. *Color Imaging Conference Proceedings.* Arizona, USA, 73–79.

Finlayson, G., Hordley, S. & Drew, M. (2002b) Removing shadows from images. *In* Heyden, A., Sparr, G., Nielsen, M. & Johansen, P. (eds.) *European Conference on Computer Vision.*

Finlayson, G., Drew, M. & Lu, C. (2009). Entropy minimization for shadow removal. *International Journal of Computer Vision, 85*(1), 35-57.

Fisher-Gewirtzman, D. & Wagner, I. A. (2003). Spatial openness as a practical metric for evaluating built-up environments. *Environment and Planning B: Planning and Design, 30*(1), 37-49.

Fisher, P. F. (1991). First experiments in viewshed uncertainty: the accuracy of the viewshed area. *Photogrammetric Engineering & Remote Sensing, 57*(10), 1321-1327.

Fisher, P. F. (1993). Algorithm and implementation uncertainty in viewshed analysis. *International Journal of Geographical Information Science, 7*(4), 331-347.

Fisher, P. F. (1994a). Stretching the viewshed. *Proceedings of the 6th International Symposium on Spatial Data Handling*, 725-738.

Fisher, P. F. (1994b) Probable and fuzzy models of the viewshed operation. *In* Worboys, M. F. (ed.) *Innovations in GIS 1.* London, UK, Taylor and Francis.

Fisher, P. F. (1995). An exploration of probable viewsheds in landscape planning. *Environment and Planning B: Planning and Design, 22*, 527-546.

Fisher, P. F. (1996). Extending the applicability of viewsheds in landscape planning. *Photogrammetric Engineering and Remote Sensing, 62*(11), 1297-1302.

Fogliaroni, P., Wallgrün, J. O., Clementini, E., Tarquini, F. & Wolter, D., (2009) A qualitative approach to localization and navigation based on visibility information. *In* Hornsby, S., Claramunt, C., Denis, M. & Ligozat, G. (eds.) *Spatial Information Theory 9th International Conference.* France, Springer, 312-329.

Förstner, W. & Gülch, E., (1987) A fast operator for detection and precise location of distinct points, corners and centres of circular features. *ISPRS Conference on Fast Processing of Photogrammetric Data.* Interlaken, 281-305.

Forsyth, D., Torr, P. & Zisserman, A. (2008) Analysis of building textures for reconstructing partially occluded facades. *In* Korah, T. & Rasmussen, C. (eds.) *Computer Vision.* Springer.

Francioni, J. M., Jackson, J. A. & Albright, L., (2002) The sounds of parallel programs. *Sixth Distributed Memory Computing Conference.* IEEE, 570-577.

Frank, C., Caduff, D. & Wuersch, M. (2004). From GIS to LBS: an intelligent mobile GIS. *IfGI Prints, 22*, 261-274.

Franklin, W. R. & Ray, C. K. (1994). Higher isn't necessarily better: Visibility algorithms and experiments. *Advances in GIS Research: Sixth International Symposium on Spatial Data Handling, 2*, 751-770.

Freeman, J. (1975). The Modelling of Spatial Relations. *Computer Graphics and Image Processing, 4*, 156-171.

Frey, M., (2007) CabBoots: Shoes with integrated guidance system. *1st International Conference on Tangible and Embedded Interaction* Louisiana, 245-246.

Fröhlich, P., Simon, R. & Baillie, L. (2009). Mobile spatial interaction. *Personal and Ubiquitous Computing, 13*(4), 251-253.

Frueh, C., Sammon, R. & Zakhor, A., (2004) Automated texture mapping of 3D city models with oblique aerial imagery. *Second international symposium on 3D data processing, visualization and transmission.* Greece, 396–403.

Fruin, S., Westerdahl, D., Sax, T., Sioutas, C. & Fine, P. M. (2008). Measurements and predictors of on-road ultrafine particle concentrations and associated pollutants in Los Angeles. *Atmospheric Environment, 42*(2), 207-219.

Gapp, K. P., (1995) Angle, distance, shape, and their relationship to projective relations. *In* J.D., M. & J.F, L. (eds.) *Proceedings of the 17th Annual Conference of the Cognitive Science Society.* July 22-25, 1995, University of Pittsburgh. Mahwah, NJ, Lawrence Erlbaum Associates, 112-117.

Garmin. (2011). What is WAAS? Retrieved 2 February 2011, from http://www8.garmin.com/aboutGPS/waas.html

Gartner, G., Cartwright, W., Peterson, M. P., Pun-Cheng, L. S. C., Mok, E. C. M., Shea, G. Y. K. & Yan, W. Y. (2007) EASYGO: A public transport query and guiding LBS. *Location Based Services and TeleCartography*. Springer Berlin Heidelberg.

Godha, S., Petovello, M. G. & Lachapelle, G. (2005). Performance analysis of MEMS IMU/HSGPS/magnetic sensor integrated system in urban canyons. *Proceedings of ION GPS-05*, 1977–1990.

Goodchild, M. F. (2007). Citizens as voluntary sensors: Spatial data infrastructure in the world of Web 2.0. *International Journal of Spatial Data Infrastructures Research, 2*, 24-32.

Goodman, J., Gray, P., Khammampad, K. & Brewster, S., (2004) Using Landmarks to Support Older People in Navigation. *6th International Symposium on Mobile Human-Computer Interaction: MobileHCI 2004.* 2004/01//, Glasgow, UK, Springer, 38-48.

Google. (2006). Take the power of Google Maps with you on your mobile phone. Retrieved 30 May 2006, from http://www.google.com/gmm/index.html

Google. (2011). City in 3D Program. Retrieved 11 February 2011, from http://sketchup.google.com/3dwh/citiesin3d/

Goose, S., Sudarsky, S., Zhang, X. & Navab, N. (2003). Speech-enabled augmented reality supporting mobile industrial maintenance. *Pervasive computing, 2*(1), 65–70.

Grice, H. P. (1975) Logic and conversation. *In* Cole, P. & Morgan, J. (eds.) *Syntax and Semantics 3: Speech Acts.* Academic Press.

Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition, , 5(2)*, 199–220.

Güsgen, H. W. (1989). Spatial reasoning based on Allen's temporal logic. *International Computer Science Institute, TR-89-049,Berkeley*.

Gulliver, J. & Briggs, D. (2007). Journey-time exposure to particulate air pollution. *Atmospheric Environment, 41*, 7195-7207.

Harris, C. & Stephens, M., (1988) A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*. Manchester, 147–151.

Hautiere, N., Aubert, D. & Dumont, E. (2007). Mobilized and mobilizable visibility distances for road visibility in fog. *26th Session of the CIE, Beijing, China*.

Hayman, E., Caputo, B., Fritz, M. & Eklundh, J.-O. (2004) On the Significance of Real-World Conditions for Material Classification. *In* Pajdla, T. & Matas, J. (eds.) *Computer Vision.* Springer.

Helvaciog lu, E. & Olguntürk, N. (2009). Colour contribution to children's wayfinding in school environments. *Optics and Laser Technology, 43*(2), 410-419.

Hernández, D. (1991) Relative representation of spatial knowledge: The 2-D case. *In* Mark, D. M. & Frank, A. U. (eds.) *Cognitive and linguistic aspects of geographic space*. Netherlands, Kluwer Academic Publishers.

Hide, C., Botterill, T. & Andreotti, M., (2010) Low cost vision-aided IMU for pedestrian navigation. *Ubiquitous Positioning Indoor Navigation and Location Based Service* Kirkkonummi, IEEE, 1-7.

Hightower, J., Consolvo, S., LaMarca, A., Smith, I. & Hughes, J., (2005) Learning and recognizing the places we go. *UbiComp 2005*. 11 September 2005, Tokyo, 159-176.

Hirtle, S. C. & Jonides, J. (1985). Evidence of hierarchies in cognitive maps. *Memory and Cognition, 13*(3), 208-217.

Hirtle, S. C. & Heidorn, P. B. (1993). The structure of cognitive maps: Representations and processes. *Behavior and Environment: Psychological and Geographical Approaches*, 170-192.

Hodge, N. (2007). US military looks to fill gaps in GPS navigation. *Jane's International Defence Review, May*.

Holland, S., Morse, D. R. & Gedenryd, H. (2002). AudioGPS: Spatial Audio Navigation with a Minimal Attention Interface. *Personal and Ubiquitous Computing, 6*(4), 253-259.

Hollerer, T., Feiner, S., Terauchi, T., Rashid, G. & Hallaway, D. (1999). Exploring MARS: Developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers and Graphics (Pergamon), 23*(6), 779-785.

Howell, M., Love, S. & Turner, M. (2005). Spatial metaphors for a speech-based mobile city guide service. *Personal and Ubiquitous Computing, 9*(1), 32-45.

Intel. (2010). OpenCV. Retrieved 5 September 2010, from http://opencv.willowgarage.com/wiki/

Ishii, H., Arita, K. & Kobayashi, M. (1993). Toward seamless collaboration media - from TeamWorkStation to clearboard. *NTT Review, 5*(1), 24-29.

Ishii, H., Kobayashi, M. & Arita, K. (1994). Iterative design of seamless collaboration media. *Communications of the ACM, 37*(8), 83-97.

Jackendoff, R. (1992). *Languages of the Mind*: Cambridge, MA: MIT Press.

Jaejun, Y. O. O., Joo, T., Park, J. H. & Lee, J. (2002). A video geographic information system for supporting bi-directional search for video data and geographic information. *Proceedings of International Symposium 2002*.

Jameson, D. & Hurvich, L. M. (1989). Essay concerning color constancy. *Annual Review of Psychology, 40*(1), 1-24.

Jiang, B. & Yao, X. (2006). Location-based services and GIS in perspective. *Computers, Environment and Urban Systems, 30*(6), 712-725.

Jianhong, G., Lu, L. & Peng, G. (2010). Removing shadows from Google Earth images. *Int. Journal of Remote Sensing, 31*(6), 1379-1389.

Kanade, T., Yoshida, A., Oda, K., Kano, H. & Tanaka, M. (1996). A stereo machine for video-rate dense depth mapping and its new applications. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 196-202.

Kang, S. B. & Szeliski, R. (1997). 3-D scene data recovery using omnidirectional multibaseline stereo. *International Journal of Computer Vision, 25*(2), 167-183.

Kanjo, E., Benford, S., Paxton, M., Chamberlain, A., Fraser, D. S., Woodgate, D., Crellin, D. & Woolard, A. (2007). MobGeoSen: facilitating personal geosensor data collection and visualization using mobile phones *Personal and Ubiquitous Computing*.

Kaplan, E. & Hegarty, C. (2006). Understanding GPS: Principles and Applications Second Edition. *Artech House*.

Kaur, S., Clark, R. D. R., Walsh, P. T., Arnold, S. J., Colvile, R. N. & Nieuwenhuijsen, M. (2006). Exposure visualisation of ultrafine particle counts in a transport microenvironment. *Atmospheric Environment, 40*(2), 386-398.

Kidner, D., Sparkes, A. & Dorey, M. (1999). GIS and wind farm planning. *Geographical Information and Planning*, 203-223.

Kim, R. (2011). Are location-based services ready to turn the corner? Retrieved 30 January 2011, from http://gigaom.com/2011/01/26/are-location-based-services-ready-to-turn-the-corner/

Kingham, S., Meaton, J., Sheard, A. & Lawrenson, O. (1998). Assessment of exposure to traffic-related fumes during the journey to work. *Transportation Research Part D-Transport and Environment, 3*(4), 271-274.

Klatzky, R. (1998) Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. *In* Freksa, C., Habel, C. & Wender, K. F. (eds.) *Spatial Cognition: An Interdisciplinary Approach to Representing and Processing Spatial Knowledge* Berlin, Springer.

Kline, B., Theresa, J., Ghali, L. M., Kline, D. W. & Brown, S. (1990). Visibility distance of highway signs among young, middle-aged, and older observers: Icons are better than text. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 32*(5), 609-619.

Kline, D. W. & Fuchs, P. (1993). The visibility of symbolic highway signs can be increased among drivers of all ages. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*(1), 25-34.

Koenderink, J. J. & Van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A, 8*(2), 377-385.

Kopf, J., Chen, B., Szeliski, R. & Cohen, M. F., (2010) Street slide: Browsing street level imagery. *ACM Transactions on Graphics.* 1-8.

Kray, C. & Porzel, R., (2000) Spatial cognition and natural language interfaces in mobile personal assistants. *Proceedings of workshop on AI in Mobile Systems.* Berlin, Germany.

Kray, C. & Kortuem, G., (2004) Interactive Positioning Based on Object Visibility. *6th International Symposium on Mobile Human-Computer Interaction: MobileHCI 2004.* Glasgow, UK, Springer, 276-287.

Krumm, J., Cermak, G. & Horvitz, E., (2003) RightSPOT: A Novel Sense of Location for a Smart Personal Object. *UbiComp 2003: 5th International Conference on Ubiquitous Computing.* Seattle, WA, USA, Springer, 36-43.

Kumler, M. P. (1994). An intensive comparison of Triangulated Irregular Networks (TINs) and Digital Elevation Models (DEMs). *Cartographica: The International Journal for Geographic Information and Geovisualization, 31*(2), 1-99.

Lachapelle, G. (2004). GNSS Indoor Location Technologies. *Journal of Global Positioning Systems, 3*(1-2), 2-11.

LaMarca, A., Chawathe, Y., Consolvo, S., Hightower, J., Smith, I., Scott, J., Sohn, T., Howard, J., Hughes, J., Potter, F., Tabert, J., Powledge, P., Borriello, G. & Schilit, B. (2005) Place Lab: Device Positioning Using Radio Beacons in the Wild. *Pervasive Computing.* Munich, Springer.

Lammens, J. M. & Shapiro, S. C., (1993) Learning symbolic names for perceived colors. *Machine Learning in Computer Vision: What, Why and How?* , AAAI Press.

Land, E. H. (1977). The Retinex Theory of Colour Vision. *Scientific American, 237*(6), 108-129.

Le Yaouanc, J.-M., Saux, É. & Claramunt, C. (2010) A visibility and spatial constraint-based approach for geopositioning. *In* Fabrikant, S., Reichenbacher, T., van Kreveld, M. & Schlieder, C. (eds.) *Geographic Information Science.* Springer.

Lee, J. (1994). Digital analysis of viewshed inclusion and topographic features on digital elevation models. *Photogrammetric Engineering & Remote Sensing, 59*, 1149-1160.

Lee, J. & Stucky, D. (1998). On applying viewshed analysis for determining least-cost paths on digital elevation models. *International Journal of Geographical Information Science, 12*(8), 891-905.

Leibe, B., Cornelis, N., Cornelis, K., Van Gool, L. & Zurich, E. T. H., (2007) Dynamic 3d scene analysis from a moving vehicle. *Conference on Computer Vision and Pattern Recognition.* Minneapolis, USA, IEEE.

Lensch, H. P. A., Heidrich, W. & Seidel, H. P., (2000) Automated texture registration and stitching for real world models. *Proceedings of the 8th Pacific Conference on Computer Graphics and Applications.* IEEE Computer Society, 317-452.

Levinson, S. C. (2003). *Space in language and cognition: Explorations in cognitive diversity*: Cambridge University Press.

Liu, L., Zhang, L., Chen, C. & Chen, H., (2008) An improved LOS method for implementing visibility analysis of 3D complex landscapes. *International Conference on Computer Science and Software Engineering.* Wuhan, China, IEEE Computer Society, 874-877.

Llobera, M. (2003). Extending GIS-based visual analysis: the concept of visualscapes. *International Journal of Geographical Information Science, 17*(1), 25-48.

Llobera, M., Wheatley, D. W., Steele, T. J. M., Cox, S. & Parchment, O. (2004). *Calculating the inherent visual structure of a landscape ('total viewshed') using high-throughput computing*. Paper presented at the XXXII International Conference - Computer Applications in Archaeology 2004 - Computer Applications and Quantitative Methods in Archaeology, Beyond the Artifact: Digital Interpretation of the Past. Retrieved. from http://eprints.soton.ac.uk/43036/

Llobera, M. (2007a). Modelling visibility through vegetation. *International Journal of Geographical Information Science, 21*(7), 799-810.

Llobera, M. (2007b). Reconstructing visual landscapes. *World Archaeology, 39*(1), 51-69.

Long, S., Aust, D., Abowd, G. & Atkeson, C., (1996) Cyberguide: prototyping context-aware mobile applications. *Conference on Human Factors in Computing Systems.* Vancouver, 293-294.

Lovelace, K. L., Hegarty, M. & Montello, D. R. (1999) Elements of good route directions in familiar and unfamiliar environments. *In* Freksa, C. & Mark, D. (eds.) *Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science.* Springer Berlin / Heidelberg.

Lucas, B. D. & Kanade, T., (1981) An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conference on Artificial Intelligence.* Vancouver, Canada, Morgan Kaufmann Publishers Inc., 674-679.

Lynch, K. (1976). *Managing the sense of a region*: MIT Press Cambridge.

Maierhofer, S., Simon, R. & Tobler, R. F., (2007) Simplified guided visibility sampling for location based services. *In* Manfred Schrenk, Vasily V. Popovich & Benedikt, J. (eds.) *REAL CORP 007.* Vienna, Austria.

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M. & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences, 8*(3), 108-114.

Malaka, R. & Zipf, A., (2000) DEEP MAP - Challenging IT research in the framework of a tourist information system. . *In* D. Klein, S. & Buhalis, D. (eds.) *Information and Communication Technologies in Tourism 2000. Proceedings of ENTER 2000.* Barcelona. Spain. , Springer Computer Science, Wien, New York. , 5-27.

Manning, C. D., Schütze, H. & MitcogNet (1999). *Foundations of statistical natural language processing*: MIT Press.

Matsuzawa, M., Takechi, H., Kajiya, Y., Ito, Y. & Igarashi, M. (2009). How drivers perceive visibility in blowing snow. *Transportation Research Record: Journal of the Transportation Research Board, 2107*(-1), 143-149.

May, A. J., Ross, T., Bayer, S. H. & Tarkiainen, M. J. (2003). Pedestrian navigation aids: information requirements and design implications. *Personal and Ubiquitous Computing, 7*(6), 331-338.

May, A. J., Ross, T. & Bayer, S. H. (2005). Incorporating landmarks in driver navigation system design: An overview of results from the REGIONAL project. *Journal of Navigation, 58*(1), 47-65.

McEllroy, J., Raquet, J. F. & Temple, M. A. (2007). Opportunistic navigation. *GPS World, 18*(7), 44-49.

Meng, L. (2005) Ego centres of mobile users and egocentric map design. *In* Meng, L., Zipf, A. & Reichenbacher, T. (eds.) *Map-based Mobile Services.* Berlin, Springer.

Merriam-Webster. (2010). Opposite [Electronic Version]. Retrieved 6 December 2010 from http://www.merriam-webster.com/dictionary/opposite.

Mezentsev, O., Lachapelle, G. & Collin, J. (2005). Pedestrian Dead Reckoning–A Solution to Navigation in GPS Signal Degraded Areas. *Geomatica, 59*(2), 175–182.

Michelis, D., Resatsch, F., Nicolai, T. & Schildhauer, T. (2008). The disappearing screen: scenarios for audible interfaces. *Personal and Ubiquitous Computing, 12*(1), 33.

Michon, P. E. & Denis, M., (2001) When and why are visual landmarks used in giving directions? *In* Montello, D. (ed.) *Spatial Information Theory: Lecture Notes in Computer Science.* Berlin, Springer, 292-305.

Microsoft Corporation. (2006). Sensors and Devices - SenseCam [Electronic Version] from http://research.microsoft.com/sendev/project_sensecam.aspx.

Micusik, B. & Kosecka, J., (2009) Piecewise planar city 3D modeling from street view panoramic sequences. *Computer Vision and Pattern Recognition.* Miami, FL, IEEE, 2906-2912.

Millonig, A. & Schechtner, K. (2007). Developing landmark-based pedestrian-navigation systems. *IEEE Transactions on Intelligent Transportation Systems, 8*(1), 43-49.

Montello, D. (1993). Scale and multiple psychologies of space. *Spatial Information Theory A Theoretical Basis for GIS,* 312-321.

Morello, E. & Ratti, C. (2009). A digital image of the city: 3D isovists in Lynch's urban analysis. *Environment and Planning B, 36,* 837-853.

Moroney, N. (2010). HP's On-line Color Thesaurus. Retrieved 11 October 2010, from http://www.hpl.hp.com/personal/Nathan_Moroney/color-thesaurus.html

Morville, P. (2005). *Ambient Findability*: O'Reilly Media Inc.

Mountain, D. & MacFarlane, A. (2007). Geographic information retrieval in a mobile environment: Evaluating the needs of mobile individuals. *Journal of Information Science, 33*(5), 515-530.

Munnich, E., Landau, B. & Dosher, B. A. (2001). Spatial language and spatial representation: A cross-linguistic comparison. *Cognition, 81*(3), 171-208.

Narzt, W., Pomberger, G., Ferscha, A., Kolb, D., Moller, R., Wieghardt, J., Hortner, H. & Lindinger, C. (2006). Augmented reality navigation systems. *Universal Access in the Information Society*, 1-11.

Neer. (2011). Neer Mobile Application. Retrieved 11 January 2011, from http://www.neerlife.com/

Nokia, (2006) Justification for the addition of carrier phase measurements. . *3GPP TGS-GERAN*. Lisbon, Portugal.

Nothegger, C., Winter, S. & Raubal, M. (2004). Computation of the Salience of Features. *Spatial Cognition and Computation, 4*(2), 113–136.

O'Donoghue, R. T., Gill, L. W., McKevitt, R. J. & Broderick, B. (2007). Exposure to hydrocarbon concentrations while commuting or exercising in Dublin. *Environment International, 33*(1), 1-8.

O'Sullivan, D. & Turner, A. (2001). Visibility graphs and landscape visibility analysis. *International Journal of Geographical Information Science, 15*(3), 221-237.

ODPM. (2006). Final report for signage and wayfinding for people with learning difficulties. Retrieved 2 Dec 2010, from http://www.communities.gov.uk/documents/planningandbuilding/pdf/144248.pdf

OGC. (2006). OpenGIS Implementation Specification for Geographic information - Simple feature access Part 1: Common architecture. 35-41.

OGC. (2008). OpenGIS Location Services (OpenLS) [Electronic Version]. Retrieved 23 November 2010 from http://www.opengeospatial.org/standards/ols.

Ohrtman, F. (2005). *WiMAX handbook: Building 802.16 wireless networks*, New York: McGraw-Hill.

Omasa, K., Hosoi, F., Uenishi, T. M., Shimizu, Y. & Akiyama, Y. (2008). Three-dimensional modeling of an urban park and trees by combined airborne and portable on-ground scanning LIDAR remote sensing. *Environmental Modeling and Assessment, 13*(4), 473-481.

OpenNLP. (2011). Open Natural Language Processing [Electronic Version]. Retrieved 3 February 2011 from http://incubator.apache.org/opennlp/.

Ott, B., Wasle, E., Weimann, F., Branco, P. & Nicole, R. (2005) Pedestrian Navigation in Difficult Environments: Results of the ESA Project SHADE. *Geo-information for Disaster Management.*

Pajdla, T., Matas, J., Hayman, E., Caputo, B., Fritz, M. & Eklundh, J.-O. (2004) On the Significance of Real-World Conditions for Material Classification. *Computer Vision - ECCV 2004.* Springer Berlin / Heidelberg.

Palmer, T. C. & Shan, J. A. (2002). Comparative study on urban visualization using LIDAR data in GIS. *URISA Journal, 14*(2), 19-25.

Patton, B. K., Aukerman, R. & Shorter, J. D. (2005). Wireless technologies, wireless fidelity (WI-FI) & worldwide interoperability for microwave access (WiMax). *Issues in Information Systems, 6*(2), 364-370.

Peuquet, D. J. (2002). *Representations of space and time*: The Guilford Press.

Polak, J. & Hoose, N. (2008). Mobile Environmental Sensing System Across Grid Environments [Electronic Version] from http://bioinf.ncl.ac.uk/message/.

Puppo, E. & Marzano, P. (1997). Discrete visibility problems and graph algorithms. *International Journal of Geographical Information Science, 11*(2), 139-161.

Pylvänäinen, T., Roimela, K., Vedantham, R., Itäranta, J., Wang, R. & Grzeszczuk, R., (2010) Automatic Alignment and Multi-View Segmentation of Street View Data using 3D Shape Priors. *3D Data Processing, Visualization and Transmission 2010.* Paris, France.

Radoczky, V. (2007) How to design a pedestrian navigation system for indoor and outdoor environments. *Location Based Services and TeleCartography.* Berlin, Springer.

Rana, S. & Morley, J. (2002). Optimising visibility analyses using topographic features on the terrain [Electronic Version]. *Centre for Advanced Spatial Analysis, University College London, London, UK*. Retrieved 23 June 2009 from http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=44.

Rana, S. (2003). Fast approximation of visibility dominance using topographic features as targets and the associated uncertainty. *Photogrammetric Engineering & Remote Sensing, 69*(8), 881-888.

Randerson, J. (2004). Photo recognition software gives location. Retrieved 24 June 2008, from http://www.newscientist.com/article/dn4857-photo-recognition-software-gives-location.html

Raper, J., Gartner, G., Karimi, H. & Rizos, C. (2007). A critical evaluation of location based services and their potential. *Journal of Location Based Services, 1*(1), 5-45.

Ratti, C., Pulselli, R. M., Williams, S. & Frenchman, D. (2006). Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design, 33*(5), 727–748.

Raubal, M. & Winter, S. (2002) Enriching wayfinding instructions with local landmarks *In* Egenhofer, M. J. & Mark, D. M. (eds.) *Second International Conference GIScience.* Boulder, USA, Springer.

Recky, M. & Leberl, F., (2010) Windows Detection Using K-means in CIE-Lab Color Space. *ICPR.* Istanbul, Turkey, 356-360.

Red Hen Systems. (2009). MediaMapper [Electronic Version] from http://www.afds.net/mediamapper.html.

RedPaw. (2008). Cathedral Square Images (12mm). Retrieved 8 August 2010, from http://photosynth.net/view.aspx?cid=a3db4c7f-f638-4860-a790-9b3b9f66814d

Reichenbacher, T. (2005) Adaptive egocentric maps for mobile users. *In* Meng, L., Zipf, A. & Reichenbacher, T. (eds.) *Map-based Mobile Services.* Berlin, Springer.

Reitberger, J., Krzystek, P. & Stilla, U. (2006). Analysis of full waveform lidar data for tree species classification. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 36*(Part 3), 228-233.

Reiter, E. & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering, 3*(1), 57-87.

Richter, K., Peters, D., Kunhnmunch, G. & Schmid, F. (2008). *What Do Focus Maps Focus On?* Paper presented at the Proceedings of the International Conference on Spatial Cognition VI. Retrieved.

Riedemann, C., (2004) Towards usable topological operators at GIS user interfaces. *In* Toppen, F. & Prastacos, P. (eds.) *Conference on Geographic Information Science (AGILE '04).* Crete, Greece, Crete University Press, 669–674.

Riggs, P. D. & Dean, D. J. (2007). An investigation into the causes of errors and inconsistencies in predicted viewsheds. *Transactions in GIS, 11*(2), 175-196.

Ross, T., May, A. & Thompson, S., (2004) The use of landmarks in pedestrian navigation instructions and the effects of context. *6th International Symposium on Mobile Human-Computer Interaction: MobileHCI 2004.* 2004/01//, Glasgow, UK, Springer, 300-304.

Rottensteiner, F. & Briese, C. (2002). A new method for building extraction in urban areas from high-resolution LiDAR data. *International Archives of Photogrammetry and Remote Sensing and Spatial Information Sciences, 34*(3A), 295-301.

Rudman, P., North, S. & Chalmers, M., (2005) Mobile Pollution Mapping in the City. *Proceedings UK-UbiNet Workshop on eScience and Ubicomp.* May 2005, Edinburgh.

Scanlan, J. M., Chabries, D. M. & Christiansen, R. W., (1990) A shadow detection and removal algorithm for 2-d images. *International Conference on Acoustics, Speech, and Signal Processing.* Albuquerque, USA, IEEE, 2057-2060.

Scharstein, D. & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision, 47*(1), 7-42.

Schlosberg, H. (1950). A note on depth perception, size constancy, and related topics. *Psychological Review, 57*(5), 314-317.

Schmid, F., Kuntzsch, C., Winter, S., Kazerani, A. & Preisig, B. (2010). *Situated Local and Global Orientation in Mobile You-Are-Here Maps*. Paper presented at the Proceedings of the 12th International Conference on Human Computer interaction with Mobile Devices and Services. Retrieved.

Shamos, M. I. (1977). Problems in computational geometry. *Unpublished manuscript, Carnegie Mellon University, Pittsburgh, PA*.

Sharma, R., Yeasin, M., Krahnstoever, N., Rauschert, I., Cai, G., Brewer, I., MacEachren, A. M. & Sengupta, K. (2003). Speech-gesture driven multimodal interfaces for Crisis Management. *Proceedings of the IEEE, 91*(9), 1327-1353.

Simon, I. & Seitz, S. (2008). Scene segmentation using the wisdom of crowds. *Computer Vision–ECCV 2008*, 541-553.

Skyhook. (2008). Location you can trust. Retrieved 14 October 2008, from http://www.skyhookwireless.com/

Slawski, B. (2007). Better business location search using OCR with street views [Electronic Version] from www.seobythesea.com/?p=701.

Snavely, N., Seitz, S. M. & Szeliski, R. (2006). Photo tourism: exploring photo collections in 3D. *ACM Transactions on Graphics, 25*(3), 835-846.

Sorrows, M. & Hirtle, S. (1999) The nature of landmarks for real and electronic spaces. *In* Freksa, C. & Mark, D. (eds.) *Spatial information theory.* Springer.

Sperber, D. & Wilson, D. (1995). *Relevance: Communication and cognition*: Oxford: Blackwell.

Spiekermann, S. (2004) General aspects of location-based services *In* Schiller, J. & Voisard, A. (eds.) *Location-based services.* Morgan Kaufmann Publishers.

Stanford, C., Williams, P., Sanderson, D., Pash, K. & Lohmeyer, D. (2003). Enhanced line-of-sight modelling and associated scenario development issues. *DSTO External Publications - SimTecT, 2003*.

Strassman, M. & Collier, C. (2004). Case study: Development of the find friend application. *Location-Based Services*, 27–40.

Stucky, J. L. D. (1998). On applying viewshed analysis for determining least-cost paths on Digital Elevation Models. *International Journal of Geographical Information Science, 12*(8), 891-905.

Sturges, J. & Whitfield, T. W. A. (1995). Locating basic colours in the Munsell space. *Color Research and Application, 20*(6), 364-376.

Sturm, P. & Triggs, B., (1996) A factorization based algorithm for multi-image projective structure and motion. *In* Buxton, B. & Cipolla, R. (eds.) *Computer Vision ECCV96.* Cambridge, UK, Springer, 709-720.

Tandy, C. R. V., (1967) The isovist method of landscape survey. *In* Murray, A. (ed.) *Symposium: Methods of Landscape Analysis.* May 1967, London, England, Landscape Research Group, 9–10.

Tarquini, F., De Felice, G., Fogliaroni, P. & Clementini, E. (2007). A qualitative model for visibility relations. *Lecture Notes in Computer Science, 4667*, 510.

Teng, Y. (1993). *Parallel processing of geometric structures: visibility and triangulation algorithms.* University of Maryland.

Terwoert, J. (2009). *EU-project VECTOR: Visualising cyclists' exposure to fine particles*. Paper presented at the Velo-City 2009. Retrieved.

Tomko, M., Trautwein, F. & Purves, R., (2009) Identification of Practically Visible Spatial Objects in Natural Environments. *AGILE 2009* Vienna, Austria, Springer-Verlag.

Triggs, B., McLauchlan, P., Hartley, R. & Fitzgibbon, A. (2000) Bundle adjustment - a modern synthesis. *In* Triggs, B., Zisserman, A. & Szeliski, R. (eds.) *Vision Algorithms: Theory and Practice*. Corfu, Greece, Springer.

Turner, A., Doxa, M., O'Sullivan, D. & Penn, A. (2001). From isovists to visibility graphs: A methodology for the analysis of architectural space. *Environment and Planning B: Planning and Design, 28*(1), 103-121.

Tversky, B., (1993) Cognitive maps, cognitive collages, and spatial mental models. *In* Frank, A. U. & Campari, I. (eds.) *Spatial Information Theory: A Theoretical Basis for GIS.* Italy, Springer-Verlag, 14-24.

US Army Corps of Engineers. (2004). Line-of-Sight Compendium [Electronic Version]. Retrieved 10 December 2007 from http://www.agc.army.mil/operations/programs/LOS/LOS%20Compendium.doc.

Van Roosbroeck, S., Wichmann, J., Janssen, N. A. H., Hoek, G., van Wijnen, J. H., Lebret, E. & Brunekreef, B. (2006). Long-term personal exposure to traffic-related air pollution among school children, a validation study. *Science of the Total Environment, 368*(2-3), 565-573.

Varges, S., (2005) Spatial descriptions as referring expressions in the MapTask domain. *In Proceedings of the 10th European Workshop On Natural Language Generation.* Aberdeen, UK.

Viethen, J. & Dale, R., (2008) The use of spatial relations in referring expression generation. *5th International Conference on Natural Language Generation.* Salt Fork OH, USA., 59-67.

Vincent, L. (2007). Taking online maps down to street level. *Computer, 40*(12), 118–120.

Wang, J., Robinson, G. J. & White, K. (1996). A fast solution to local viewshed computation using grid-based digital elevation models. *Photogrammetric Engineering and Remote Sensing, 62*(10), 1157-1164.

Wang, J. & Li, C., (2008) Acquisition of UAV images and the application in 3D City modelling. *In* Zhou, L. (ed.) *International Symposium on Photoelectronic Detection and Imaging* Beijing, China.

Wang, Q. & You, S., (2010) Automatic Registration of Large-Scale Multi-sensor Datasets. *11th European Conference on Computer Vision (ECCV).* September 5-11, Greece.

Weiser, M. (1993). Some Computer Science Issues in Ubiquitous Computing. *Communications of the ACM, 36*(7), pp. 75-84.

Weiser, M. (1994). The World is not a desktop. *Interactions, January 1994*, pp. 7-8. .

Weiser, M. & Brown, J. S. (1996). Designing calm technology. *PowerGrid Journal, 1.01*.

Weiser, M., Gold, R. & Brown, J. S. (1999). Origins of ubiquitous computing research at PARC in the late 1980s. *IBM Systems Journal, 38*(4), 693-695.

Werner, S., Krieg-Bruckner, B., Mallot, H. A., Schweizer, K. & Freksa, C. (1997) Spatial cognition: The role of landmark, route, and survey knowledge in human and robot navigation. *In* Jarke, M. (ed.) *Informatik '97 GI Jahrestagung.* Berlin, Heidelberg, New York. Springer.

Wheatley, D. (1995) Cumulative viewshed analysis: a GIS-based method for investigating intervisibility, and its archaeological application. *In* Lock, G. & Stancic, Z. (eds.) *Archaeology and Geographical Information Systems: A European Perspective.* Taylor & Francis.

Williams, I. (2009). Gartner predicts dramatic rise in LBS. Retrieved 19 Janurary 2010, from http://www.v3.co.uk/v3/news/2245575/location-services-rise

Winter, S. (2003) Route adaptive selection of salient features. *In* Kuhn, W., Worboys, M. F. & Timpf, S. (eds.) *Spatial Information Theory*. Springer.

Winter, S., Tomko, M., Elias, B. & Sester, M. (2008). Landmark hierarchies in context. *Environment and Planning B: Planning and Design 35*(3), 381 – 398.

Winter, S. & Wu, Y., (2008) The "spatial Turing test". *In* Navratil, G. (ed.) *Colloquium for Andrew U. Frank's 60th Birthday. Geoinfo Series, Vienna, Austria, Department for Geoinformation and Cartography*. Technical University Vienna, 109–116.

Winter, S. & Wu, Y. (2009) Intelligent Spatial Communication. *In* Navratil, G. (ed.) *Research Trends in GIS.* Berlin, Springer.

Wirola, L., Halivaara, I., Verhagen, S. & Tiberius, C. (2007). On the feasibility of adding carrier phase–assistance to cellular GNSS assistance standards. *Journal of Global Positioning Systems, 6*(1), 1-12.

Witte, T. H. & Wilson, A. M. (2004). Accuracy of non-differential GPS for the determination of speed over ground. *Journal of Biomechanics, 37*(12), 1891-1898.

Worboys, M. F. (2001). *GIS: A Computing Perspective,* London: Taylor Francis.

Wurm, L. H., Legge, G. E., Isenberg, L. M. & Luebker, A. (1993). Color improves object recognition in normal and low vision. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 899-911.

Xiao, Q., Ustin, S. L. & McPherson, E. G. (2004). Using AVIRIS data and multiple-masking techniques to map urban forest tree species. *International Journal of Remote Sensing, 25*(24), 5637-5654.

Yang, P. P. J., Putra, S. Y. & Li, W. (2007). Viewsphere: A GIS-based 3D visibility analysis for urban design evaluation. *Environment and Planning B: Planning and Design, 34*(6), 971-992.

Yang, Z. S., Wang, W., Dong, S., Zhu, W. Q. & Shen, J. H. (2008). Information fusion technology of GPS/DR integrated positioning system. *Journal of Jilin University (Engineering and Technology Edition), 38*(3), 508-513.

Yell. (2010). Explore Famous UK Places from Above [Electronic Version]. Retrieved 16 August 2010 from http://www.yell.com/maps/MapAction.do.

Yi-Leh, W., Cheng-Yuan, T., Maw-Kae, H. & Chi-Tsung, L. (2010). Automatic image interpolation using homography. *EURASIP Journal on Advances in Signal Processing, 2010*, 1-12.

Ying, S., Li, L., Mei, Y. & Peng, X., (2006) Incremental terrain visibility analysis. *Proceedings of SPIE - The International Society for Optical Engineering.* 28 October 2006, Wuhan, SPIE.

Yu, S., Al-Jadir, L. & Spaccapietra, S., (2005) Matching User's Semantics with Data Semantics in Location-Based Services. *1st Workshop on Semantics in Mobile Environments* Cyprus.

Zeiner, H., Kienast, G., Derler, C. & Haas, W. (2005). Video documentation of urban areas. *Computers, Environment and Urban Systems, 29*(6), 653-668.

Zhou, J., Chu, K. M. K. & Ng, J. K. Y. (2005). Providing location services within a radio cellular network using ellipse propagation model. *International Conference on Advanced Information Networking and Applications, 1*, 559-564.

Zipf, A. (2004). An Evaluation of the OpenLS Specifications for multi-modal mobile applications. *Journal of Geographic Information Sciences, 10*(2), 117-127.

Zubin, D. (1989) Natural language understanding and reference frames. *In* Mark, D., Frank, A., Egenhofer, M. J., Freundschuh, S., McGranaghan, M. & White, R. (eds.) *Languages of Spatial Relations: Initiative Two Specialist Meeting Report.* Santa Barbara, CA, National Center for Geographic Information and Analysis.

Zwahlen, H. T. & Schnell, T. (1999). Visibility of road markings as a function of age, retroreflectivity under low-beam and high-beam illumination at night. *Transportation Research Record: Journal of the Transportation Research Board, 1692*(-1), 152-163.